ANALYZING THE POLICY IMPLICATIONS OF SUBSIDIES, CAUSALITY STRUCTURES, AND PRE-DETERMINED DEMAND ON CONSUMER FOOD ACQUISITIONS

A Dissertation

by

MARK CHRISTOPHER SENIA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Oral Capps, Jr.
Co-Chair of Committee,	Senarath Dharmasena
Committee Members,	Ariun Ishdorj
	Marco A. Palma
Head of Department,	C. Parr Rosson III

August 2017

Major Subject: Agricultural Economics

ABSTRACT

The overall objective of this dissertation is to contribute to a better understanding of consumer food acquisitions by considering policies to promote dietary fiber intake, modeling consumer food acquisitions as a complex system, and by determining the effects of pre-determined demand and regularity conditions on policy analysis. To achieve these objectives, three related empirical investigations of consumer food acquisitions are conducted.

The first paper conducts a panel regression on nine per-capita fiber intake categories taken from purchases of a variety of food to uncover socioeconomic, demographic and government food policy related factors on the per capita intake of dietary fiber in the United States. Although consumer response to 2010 *Dietary Guidelines for Americans* in terms of increased intake of dietary fiber showed mixed results, a proposed 20% subsidy on prices of fruits and vegetables showed some promising results concerning increasing fiber intake in the U.S. diet.

The second uses individual and household attributes, characteristics of the local food environment, the individual's dietary pattern, prices, health outcomes, and policy variables jointly to estimate a complex graphical causality structure. The resulting directed acyclic graph shows a number of complicated relationship among these variables. Concerning the paths between poverty, race and food insecurity, we find a number of paths. Thus, policymakers that want to reduce the problems associated with food insecurity need a full picture of the complex interactions among all these variables.

In addition, we find variables associated with the Supplemental Nutrition Assistance Program participation and food insecurity to be not strictly endogenous. Obesity was found to be strictly endogenous.

The objective of the third paper is to examine the affect that ignoring pre-determined demand and theoretical regularity conditions will have on consumer food demand. To accomplish this we used the Almost Ideal Demand System because of its wide use in applied policy research. A major result from this study is that elasticities calculated under the presence of pre-commitments are more elastic relative to those calculated without. The result from a proposed subsidy further reinforces the importance of accounting for pre-commitments. In terms of satisfying regularity conditions, the AIDS with pre-commitments performs slightly better. One further important result from this study is not only the need to account for pre-commitments, but also the need to account for the timing of a consumer's pre-commitments, since pre-committed quantities could vary over time.

DEDICATION

To my wife and family.

ACKNOWLEDGEMENTS

I would like to thank Senarath Dharmasena for his tireless guidance and support throughout the course of this research. I would also like to thank Oral Capps, Jr., Ariun Ishdorj, and Marco A. Palma for being a part of the committee.

Thanks also go to David J. Leatham and the Department of Agricultural Economics for giving me to opportunity to further my education. I appreciate the support given to me by the department and faculty in my graduate career.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Senarath Dharmasena as advisor, Dr. Oral Capps, Jr., Dr. Ariun Ishdorj, and Dr. Marco A. Palma of the Department of Agricultural Economics.

All work for the dissertation was completed by the student, in collaboration with Senarath Dharmasena of the Department of Agricultural Economics.

Funding Sources

Graduate study was supported by the Robert G. Cherry Graduate Fellowship and a teaching assistantship from the Department of Agricultural Economics at Texas A&M University.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

This dissertation examines individuals' dietary choices, acquisitions, and interactions within the food environment. These behaviors are critically important in determining an individual's dietary quality and risk of negative health outcomes such as obesity. Obesity is a major risk factor for diabetes, cardiovascular disease, cancer, sleep apnea, nonalcoholic fatty liver disease, osteoarthritis, and other problems (Ahima and Lazar, 2013). It is important for policymakers concerned with food intake and nutrition to have a full picture of the interactions among all variables.

The overall objective of this dissertation is to contribute to a better understanding of consumer food acquisitions by: (1) considering policies to promote dietary fiber intake; (2) modeling consumer food acquisitions as a complex economic system; and (3) determining the effects of pre-determined demand for food products and theoretical regularity conditions on policy analysis. To achieve these objectives, three related empirical investigations of consumer food acquisitions are conducted (Chapters II-IV).

The first paper contributes to the literature by conducting panel regressions on nine per-capita fiber intake categories taken from purchases of a variety of food types such as bread, pasta, tortilla, fresh fruit, fresh vegetables and beans, frozen fruit, frozen vegetables and beans, canned fruit, and canned vegetables and beans. The results are used to uncover socioeconomic and government food policy related factors on the per capita intake of dietary fiber in the United States. Understanding the factors influencing consumers' intake of fiber and whether dietary guidelines have an effect on this intake is an important issue for food processors as well as government policy makers to make strategic decisions. The results will be used to determine whether the recent U.S. dietary guidelines have had any effect on fiber intake. Further, the results will be used to determine the effect on dietary fiber intake from four different scenarios of a 20% price subsidy on canned, fresh, and frozen fruits and vegetables. In addition, the results are used to find a subsidy necessary to meet the guideline.

The contribution of the second study is to use the individual and household attributes, characteristics of the local food environment, the individual's dietary pattern, prices, health outcomes, and policy variables jointly to estimate a complex graphical causality structure. This is in contrast to studies that consider these variables in a fragmented approach and not as a single complex system. The estimation of a graphical causal structure is accomplished using two machine-learning algorithms: Greedy Equivalence Search (GES) and Linear non-Gaussian Orientation Fixed Structure Rule Three (LOFS R3). First, the GES algorithms is run on the data to build a graphical causality structure. Then, the LOFS R3 algorithm is run on the resulting structure to orient any edges that were not oriented by the GES algorithm. The Directed Acyclic Graph (DAG) is generated under assumptions made by imposing a priori knowledge on the structure. Then we will calculate parameter estimates underlying the structural relationships built from the causality structure of the variables. Finally, we will make comparisons of the causality effects from the estimated directed acyclic graph and parameter estimates to current research in the field.

The objective of the third paper is to examine the impact that ignoring pre-determined demand and theoretical regularity conditions will have on consumer food demand. To accomplish this, we use the Almost Ideal Demand System model because of its wide use in applied policy research. We pay additional attention to regularity by testing for compliance with these conditions. We perform the empirical analysis using Nielsen Homescan data. We create a monthly time series of a representative U.S. consumer's purchases of fresh fruit, fresh vegetables and beans, frozen fruit, frozen vegetables and beans, canned fruit, canned vegetables and beans for the years 2004 through 2014. This research estimates the presence and levels of pre-committed demand. If pre-committed demand is present, then models that do not account for this are incorrectly specified. Further, the results are used to determine the effect of a 20% price subsidy on canned, fresh, and frozen fruits and vegetables.

CHAPTER II

GUIDELINES OR SUBSIDIES: PROMOTING DIETARY FIBER INTAKE THROUGH POLICY IN THE UNITED STATES

The *2015-2020 Dietary Guidelines for Americans* recommend that individuals have an intake of 14 grams of dietary fiber per 1,000 kcal consumed per day (about 25 grams per day for a 2,000 calorie diet) (HHS and USDA, 2015). The *Dietary Guidelines for Americans* are the main source of dietary recommendations for health professionals and government agencies in the United States and are published roughly every five years. Dietary fiber is considered a nutrient of public health concern because increases the intake of fiber can reduce chronic disease risk (HHS and USDA, 2015, p. 60). One important goal of these guidelines is to increase the consumption of foods high in fiber. Examples of these foods include fruits and vegetables, beans, whole grains, and nuts. Yet despite encouragement from the government, consumers in the United States do not purchase enough foods high in dietary fiber. The average daily per capita dietary fiber intake in the United States was 16 grams per day in 2009-2010 (Hoy and Goldman, 2010). One example where consumers fall short is the purchase of few whole grain products (high in dietary fiber) relative to many refined grain products (low in dietary fiber) (Volpe and Okrent, 2013).

Dietary fiber provides a range of important health benefits particularly in preventing heart disease and diabetes. Clinical research finds that the intake of dietary fiber from cereals and fruits are inversely associated with the risk of coronary heart disease (Pereira et al., 2004). Furthermore, increased intake of dietary fiber may reduce cardiovascular disease, stroke, hypertension, diabetes, obesity, and some gastrointestinal diseases (McKeown et al., 2002; Anderson et al., 2009). There may be an association between adults who eat more whole grains, particularly those high in dietary fiber, and a lower body weight relative to adults who eat fewer whole grains (USDA, 2010). These health benefits from the dietary fiber make it an important dietary component.

The government can influence its citizens' diets in a number of methods. One option is to publish guidelines for a healthy diet formulated through recommendations from nutrition and dietary experts. The *Dietary Guidelines for Americans* are a major example. These guidelines have been shown to lead to a decrease in intake of calories derived from consumption of nonalcoholic beverages (Dharmasena, et al, 2011) and an increase in the demand for whole grain products (Mancino and Kuchler, 2012). It may then be expected that the guidelines can influence others areas of the diet such as increasing the intake of calcium, potassium, dietary fiber, and vitamin D.[1]

Another option available for the government is to influence the price of a product to encourage more consumption of this product. For example, it is possible that a 10% subsidy for low-income Americans could increase their consumption of fruits by 2.1-5.2% and vegetables by 2.1-4.9% (Dong and Lin, 2009). A 20% subsidy on healthy dishes in a university cafeteria was followed by a 6% increase in the consumption of healthy foods and a 2% decline in the consumption of less-healthy foods (Michels et al.,

[1] The *2015-2020 Dietary Guidelines for Americans* define these nutrients as under consumed and as nutrients of public health concern because low intakes are associated with health concerns (HHS and USDA, 2015, p. 60).

2008). Experiments in laboratory settings have demonstrated that a reduction in the price of certain healthier products by 10% led to an increase in the purchase of these products by 10.3% (Epstein et al., 2010).

This paper contributes to the literature by conducting panel Tobit regressions on per-capita fiber intake derived from nine categories of food products. These are taken from the purchases of a variety of food types including bread, pasta, tortilla, fresh fruit, fresh vegetables and beans, frozen fruit, frozen vegetables and beans, canned fruit, and canned vegetables and beans. The panel regression is used to uncover socioeconomic and government food policy related factors on the per capita intake of dietary fiber in the United States. Understanding the factors influencing consumers' demand for fiber and whether dietary guidelines have an effect on this demand is an important issue for food processors as well as government policy makers. The results will be used to determine whether the recent U.S. dietary guidelines have had any effect on fiber intake. Further, the results will be used to determine the effect on dietary fiber intake from four different scenarios of a 20% subsidy on canned, fresh, and frozen fruits and vegetables.

We perform the empirical analysis using the Nielsen Homescan Consumer Panel. We create a quarterly panel of the same households participating for the years 2004 through 2014. This dataset is well suited to the analysis as information is collected on purchases from participating panelists. Also, the dataset provides a wealth of socioeconomic and demographic information pertaining to each household.

Our main findings can be briefly summarized as follows: Those living below 130% and between 130% and 185% of the poverty level seem to purchase significantly less fiber per capita relative to those living above these poverty levels. For the mean household in the sample there is a 5.5% decrease in per capita dietary fiber purchase in the time period after the dietary guidelines were released compared to before the release in 2010. If the objective of these guidelines was to increase the intake of dietary fiber then this particular objective has not been fulfilled.

Regional effects in fiber purchases are also evident in that the Northeast region purchases a larger amount of fiber from bread and pasta relative to the South. We find that the own-price elasticities with regards to fiber intake for all categories show fiber intake to be inelastic. Our estimates of the own-price elasticities for fiber from fruit range from -0.25 for canned fruit to -0.55 for frozen fruit. Our estimates for the own-price elasticities for fiber from vegetables range from -0.17 for frozen vegetables to -0.33 for canned vegetables. The results also indicate that the various forms of vegetable and fruit fiber are likely not substitutes for each other. A proposed 20% subsidy applied to all categories of fruits and vegetables would result in an increase in the average per capita intake of fiber per day by 4.8%. In addition, a whopping subsidy of 2755% applied to only fresh fruits and vegetables would be necessary to meet a guideline of 25 grams of fiber per capita per day. Thus, subsidies alone would not be easily able to encourage consumers to meet the daily fiber intake guideline.

The remainder of this paper proceeds as follows. In the Literature Review section, we discuss the existing literature on fiber intake and the purchase of products high in fiber. In the Empirical Model and Estimation Procedure section, we specify the econometric model and outline the estimation methodology. In the Data section, we give a detailed description of the data and the constructed dependent and explanatory variables. In the Results section, we discuss and present the results. In Conclusion, Implications, and Limitation section, we summarize the results and discuss relevance.

Literature Review

The average per capita U.S. dietary fiber intake of 16 grams per day in 2009-2010 (Hoy and Goldman, 2010) falls far short of the average recommendation of 25 grams per day. A lack of availability is not the problem. The U.S. food supply has 25 grams per capita per day of dietary fiber available for each citizen (USDA, ERS, 2015). The majority of this availability is from grains (35.1%), vegetables (22.7%), legumes, nuts, and soy (16.2%), and fruits (11.3%). By failing to intake a sufficient amount of dietary fiber, Americans cannot enjoy the range of important health benefits provided by its intake. Most important are its possible associations with a reduction of heart disease and type 2 diabetes.

According to one clinical study, the intake of dietary fiber from cereals and fruits is inversely associated with the risk of coronary heart disease (Pereira et al., 2004). This reduction may be as high as a 40% lower risk of coronary heart disease (Rimm et al., 1996). A high intake of dietary fiber is also associated with a lower risk of metabolic syndrome, a set of medical conditions that increase the risk of developing heart disease and diabetes (McKeown et al., 2002).

Dietary fiber also appears to be an important component in lowering the risk of developing type 2 diabetes. It is thought that a diet high in dietary fiber and lower in high-glycemic-index foods may be associated with a reduction in developing diabetes for men (Fung et al., 2002) and for women (Liu et al., 2000). More specifically, certain communities, such as U.S. women that are of African origin, face a much higher incidence of type 2 diabetes and demonstrate the need for increasing dietary fiber intake (Krishnan et al., 2007).

Increased intake of dietary fiber may also reduce stroke, hypertension, obesity, and some gastrointestinal diseases (McKeown et al., 2002; Anderson et al., 2009). There may be an association between adults who eat more whole grains, particularly those high in dietary fiber, and a lower body weight relative to adults who eat fewer whole grains (USDA, 2010). These important health benefits from dietary fiber intake make it a vital component of the U.S. diet.

Current literature dealing solely with consumer dietary fiber intake demand is limited. Miguel and Diansheng (2012) use a dynamic Tobit model that allows past purchase occasions to affect current purchase decisions for fiber using the Nielsen Homescan Consumer Panel. The authors find that participation in the Supplemental Nutrition Assistance Program for Women, Infants and Children (WIC program), one of U.S. federal government's food and nutrition assistance program, the age and presence of children between thirteen and seventeen, not being Hispanic, and the employment level of the female head do not significantly affect fiber intake. Also the authors find that the female head's education level has a negative impact on fiber purchases and coupon use has a positive effect. The authors do not include fiber from fresh or frozen fruits and vegetables and do not separate the sources of dietary fiber into separate food categories.

The effect of nutritional information on nutrient intake is a popular closely related line of research. Variyam, Blaylock, and Smallwood (1996) conducted a survey on the fiber content of food and attitudes toward consumption of foods high in fiber. The authors find that knowledge of nutritional information has an influence on fiber intake. According to this study, the major factors affecting fiber intake are household income, meal planner age, smoking status, vegetarian status, race, and ethnicity. Education exerts a sizable effect by enhancing the information level. Ollberding, Wolf, and Contento (2011) use the 2005-2006 National Health and Nutrition Examination Survey data to find that food label users report higher fiber intake than those that do not use food labels in making food purchase decisions. Thus it is likely that in our sample, higher-educated individuals will have higher fiber intake.

The extant literature has previously examined the impact of the 1994 Nutrition Labeling and Education Act of the United States. Variyam (2008) examined the impact of thirteen nutrients on consumer diets displayed on the consumer nutrition label. According to this study, when consumers use nutrition labels, they increase their fiber intake by 0.69 grams per 1000 calories. Using the same data and a different estimation technique, Kim, Nayga and Capps (2000) reported that consumer nutrition label use increased the average daily fiber intake of consumers by 7.51 grams.

The literature has also focused on whole grain products (a good source of dietary fiber) likely due to the USDA making specific whole grain intake recommendations in 2005. Mancino et al. (2008) found that the release of 2005 Dietary Guidelines for Americans increased the availability and sales of whole-grain foods, with a large impact due to reformulation of existing products. Lin and Yen (2008) use the 1994–1996 Continuing Survey of Food Intakes by Individuals (CSFII) to examine how nutrition knowledge and socio-demographic variables affect the consumption of refined and whole grain products. Mancino and Kuchler (2012) estimated the demand for whole grain bread to determine if the release of the 2005 Dietary Guidelines for Americans affected demand for whole grain. They found an increase in demand for whole grain products even after accounting for price changes.

It is important for policymakers concerned with food intake and nutrition to know if their guidelines are effective in changing the behavior of citizens. Thus, testing whether or not government regulations are effective in changing behavior is an important area of research. Palma and Jetter (2012) observed no significant changes in the consumption of major food groups (e.g., meat, fruit, vegetables, etc.) over the period 2000 to 2009. This result coincides with the release of the 2005 guidelines, and we expect similar results for the 2010 guidelines without any major food policy changes.

In addition to publishing dietary guidelines, governments also can attempt to make a desired food group more widely available by allowing a consumer to pay less, otherwise known as subsidizing a food group. Comparable research does estimate the demand for food rich in fiber. Dong and Lin (2009) estimate that a 10% subsidy would encourage low-income Americans to increase their consumption of fruits by 2.1-5.2% and vegetables by 2.1-4.9%. Klerman, Bartlett, Wilde, and Olsho (2014) studied the effects of the USDA Healthy Incentives Pilot, which provided a 30% incentive for purchases of certain fruits and vegetables. These authors found that participants had a

24% higher intake of these fruits and vegetables compared to those in the control group. Lin, Yen, Dong, and Smallwood (2010) find that a 10% price subsidy for Supplemental Nutritional Assistance Program (SNAP) recipients focused on fruits and vegetables increased at-home consumption of vegetables from 0.94 to 1 per cup (6% increase) and fruits from 0.38 to 0.42 per cup (11% increase).

Nordström and Thunström (2011) estimated that a 50% subsidy on Keyhole-labelled (a symbol of healthier foods controlled by the Swedish National Food Administration) bread and breakfast cereals would lead to a 35% increase in the intake of fiber. Waterlander et al. (2012) used a sample in the Netherlands and conducted an online experiment on shopping behavior. The authors found that a 25% discount on the total amount of fruit and vegetables purchased would lead to a 25% increase fruits and vegetables purchase. Nnoaham et al. (2009) estimated that for a United Kingdom sample that a 17.5% subsidy along with a tax on less healthy food would lead to a 5% increase in fruit and vegetable consumption. Another experiment showed that a 20% subsidy on healthy dishes in a university cafeteria was followed by a 6% increase in the consumption of healthy foods and a 2% decline in the consumption of less-healthy foods (Michels et al., 2008). Experiments in laboratory settings have demonstrated that a reduction the price of certain healthier products by 10% led to an increase in the purchase of these products by 10.3% (Epstein et al., 2010).

Studies have also shown that revisions to government food assistance programs can lead to changes in diet. Andreyeva and Luedicke (2013) found that the 2009 WIC revisions increased the share of whole-grain bread and brown rice purchased while not increasing the total amount purchased. WIC households used their benefits to change some of their bread purchases, rather than to buy more bread overall, whereas total rice purchases increased.

As such, by investigating previous literature on U.S. consumers' dietary fiber intake, it is clear that only a limited amount of research has been conducted incorporating the demand for dietary fiber. Therefore, a more comprehensive study incorporating dietary fiber intake derived from various food types, the impact of socio-economic-demographic factors, and the role of government nutrition education programs promoting fiber intake is needed. This lack of information in the extant literature also warrants our study.

Empirical Model and Estimation Procedure

It is clear that not all households (or individuals) may have purchased all food types during the sampling period, resulting in zero expenditure levels (and quantities) reported for some foods under consideration. This is known as *censoring* in data. The application of ordinary least squares (OLS) to estimate a regression with a censored dependent variable can result in biased estimates, even asymptotically (Kennedy, 2003). Therefore, to account for zero instances (or censored data) of fiber intake, we adopt a Tobit model (Tobin, 1958; Amemiya, 1984). To account for the panel nature of the data, a random effects panel Tobit model is used to estimate each fiber demand (Maddala, 1987). This means that the unobservable factors that differentiate individuals in the panel are assumed to be randomly distributed variables. The individuals in the panel are likely to differ in culture, tastes, and other unobservable factors. Thus, it is reasonable to assume that the differences between them are randomly distributed.

Let y_{it}^* (e.g., dietary fiber intake from bread) be a continuous latent variable described by the following equation with panel-level random effects:

$$y_{it}^* = \boldsymbol{x}_{it}\,\beta + v_i + \varepsilon_{it}, \tag{2.1}$$

for *i=1,2,…,N, t=1,2,…,T*. For equation (2.1), $\boldsymbol{x}_{it}$ is the vector of explanatory variables for individual *i* (e.g., prices and demographics), β is the vector of coefficients to estimate, v_i are the random effects which are independent and identically distributed (i.i.d.) $\mathcal{N}(0, \sigma_v^2)$, and ε_{it} are the error terms which are i.i.d. $\mathcal{N}(0, \sigma_e^2)$ and independent of v_i. To account for the censored nature of the purchase data, we specify that the observed purchase in activity *j*, y_{ij}, is related to the latent variable y_{ij}^* as follows:

$$y_{ij} = \begin{cases} y_{ij}^* = \boldsymbol{x}_{ij}\,\beta + v_i + \varepsilon_{ij}, & if\ \ y_{ij}^* > 0 \\ 0, & if\ \ y_{ij}^* \le 0 \end{cases}, \tag{2.2}$$

where the time subscript has been dropped.

We estimate the model parameters using the Stata XTTOBIT command (StataCorp, 2015). This estimation approach is sensitive to outliers. To achieve quicker convergence and results that are more reliable the dependent variables are capped at four standard deviations above the mean value during the estimation stage. To allow interpretation of the results, we calculate and report marginal effects associated with the explanatory variables for each of the fiber intake categories (Greene, 2012, p. 848-850; McDonald and Moffitt, 1980). We further calculate elasticities for the price variables (calculation of price variables is explained in the Data section below). The unconditional marginal effects are defined as

$$ME_{x_{ij}} = \beta_{ij}\, \Phi\left(\frac{\overline{x_{ij}}\,\beta}{\sigma}\right), \tag{2.3}$$

and the conditional marginal effects are defined as

$$ME^{*}_{x_{ij}} = \beta_{ij}\left[1 - \frac{\overline{x_{ij}}\,\beta}{\sigma}\frac{\phi\left(\frac{\overline{x_{ij}}\,\beta}{\sigma}\right)}{\Phi\left(\frac{\overline{x_{ij}}\,\beta}{\sigma}\right)} - \frac{\phi\left(\frac{\overline{x_{ij}}\,\beta}{\sigma}\right)^{2}}{\Phi\left(\frac{\overline{x_{ij}}\,\beta}{\sigma}\right)^{2}}\right]. \tag{2.4}$$

Given quantities (unconditional average quantity $\overline{y_{ij}}$ and conditional average quantity $\overline{y_{ij}}^{*}$) in level form and prices in logarithmic form, the unconditional elasticities, $\epsilon_{x_{ij}}$ are defined as

$$\epsilon_{x_{ij}} = \frac{\beta_{ij}}{\overline{y_{ij}}}\, \Phi\left(\frac{\overline{x_{ij}}\,\beta}{\sigma}\right), \tag{2.5}$$

and the conditional elasticities, $\epsilon^{*}_{x_{ij}}$ are defined as

$$\epsilon^{*}_{x_{ij}} = \frac{\beta_{ij}}{\overline{y_{ij}}^{*}}\left[1 - \frac{\overline{x_{ij}}\,\beta}{\sigma}\frac{\phi\left(\frac{\overline{x_{ij}}\,\beta}{\sigma}\right)}{\Phi\left(\frac{\overline{x_{ij}}\,\beta}{\sigma}\right)} - \frac{\phi\left(\frac{\overline{x_{ij}}\,\beta}{\sigma}\right)^{2}}{\Phi\left(\frac{\overline{x_{ij}}\,\beta}{\sigma}\right)^{2}}\right], \tag{2.6}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function (CDF), $\phi(\cdot)$ is the standard normal probability density function (pdf), $\overline{x_{ij}}$ is the mean value of the dependent variable and standard error of the regression σ is defined as $\sigma = \sqrt{\sigma_{v}^{2} + \sigma_{e}^{2}}$.

Care must be taken in the interpretation of the elasticity measures presented in this paper and in particular, the conditional elasticity measure (equation 2.6) as it is the focus of the results section. Since our quantity variable is dietary fiber intake, our elasticities will be own-price elasticities of demand for fiber intake. For the conditional elasticities using bread as an example, we have the own price elasticity of demand for fiber intake from bread conditional on a positive intake of fiber from bread.

In order to find the effect of a proposed subsidy we begin by finding a baseline intake of dietary fiber as an average of the last four quarters of the data for each household. Then for each dietary fiber intake category, we increase or decrease this baseline amount by the corresponding conditional own- and cross-price elasticities. This procedure assumes that any increase in fiber demand will be met by an increase in supply at the current price. This is not as bold an assumption as it appears since the current intake of dietary fiber is 16 grams per day versus a supply of 25 grams per capita per day in the United States (Hoy and Goldman, 2010; USDA, ERS, 2015). This is a situation with a relatively inelastic demand curve and a perfectly elastic supply curve, which means a 100% pass through of the price reduction to consumers. Four different scenarios are analyzed for the 20% subsidy: on all fruit and vegetables, only canned fruit and vegetables, only fresh fruit and vegetables, and only frozen fruit and vegetables.

Data

Data are obtained from Nielsen Homescan Consumer Panel.[2] We create a quarterly panel of households for the years 2004 through 2014 (44 quarters) consisting of 9,896 households across the United Sates totaling 435,424 observations. Creating the panel in this manner leaves an active sample in which 94% of households report a positive intake of dietary fiber in either 43 or 44 of the 44 quarters. For the Nielsen

[2] Data from The Nielsen Company (US), LLC and marketing databases provided by the Kilts Center for Marketing Center at The University of Chicago Booth School of Business.

Homescan Consumer Panel, each participating household is given a scanner to read UPCs off products purchased at stores. Nielsen matches the scanned UPC with product characteristics in their database. The household is also asked to enter quantity, expenditure, and any coupon information about the products purchased.

For a selection of fiber rich food products the quantity of each food product and fiber quantities derived from these products and demographic characteristics of the household are used. The food products selected for study are bread, pasta, tortilla, fresh fruit, fresh vegetables and beans, frozen fruit, frozen vegetables and beans, canned fruit, canned vegetables and beans. Each product in the dataset is recorded with its associated Universal Product Code (UPC) number and an abbreviated product description. For each product, an estimate is made of the fiber content by utilizing keyword search over the abbreviated product descriptions. Appendix A gives details on the assumptions and gives reference numbers to the USDA National Nutrient Database for Standard Reference. Appendix B gives an example of the 10 most common item abbreviations and their assumed meanings. We are able to identify around 154,000 products across the aforementioned nine food categories. Then the fiber content for each category is summed to create the total fiber intake for the household in that quarter for each of the nine categories. This total for each category is then divided by the number of members of the household to create an approximation of daily dietary fiber intake per capita.[3]

[3] It is not a guarantee that the total amount of food purchased will be consumed. Thus, this number will represent the total maximum possible daily rate of dietary fiber intake.

Table 2.1 lists summary statistics of the dependent variables. The means presented are conditional on a positive intake in that category. The largest sources of fiber based on the conditional means are bread, canned vegetables, and fresh vegetables. It is important to notice the large number of zero observations for some of the categories. Appendix C provides a closer look at the number of quarters each household reports a greater than zero total daily fiber per capita intake. A few households report unusually large fiber purchases. This may be due to reporting issues, problems estimating the fiber content of certain foods, or the household purchasing food for members outside of the household (donations to food banks as one possible example).

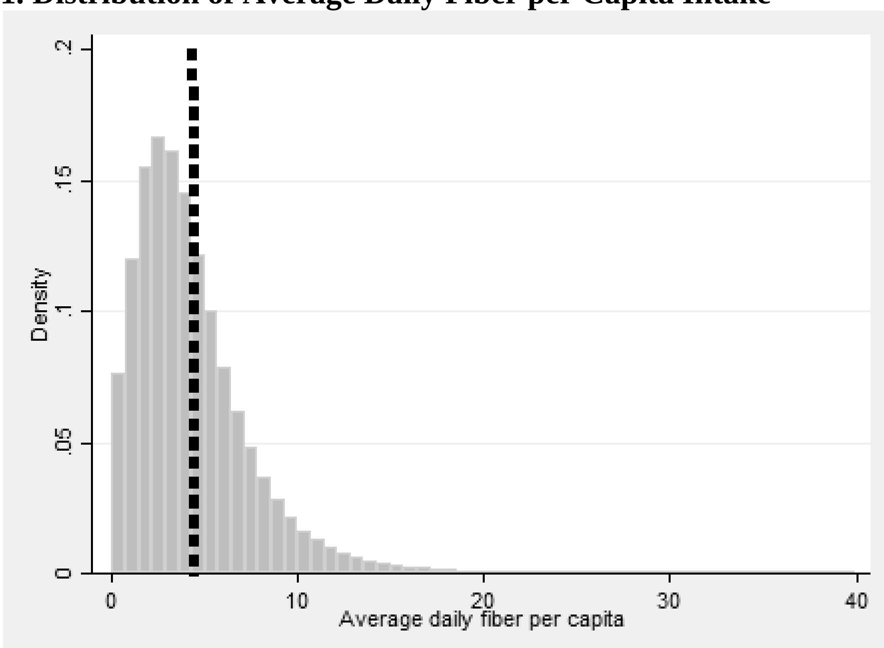
Figure 2.1 shows a histogram for the daily fiber intake per capita. The black dashed vertical line at 4.38 grams per capita per day represents average daily fiber intake per capita of our sample. This falls far short of a USDA target of 25 grams per capita per day. The majority of our sample is not meeting the USDA guidelines. The USDA (2010, pg. 46) estimates the typical American diet provides 40% of needed fiber. Our sample average shows participants meeting 16% of the recommendation. This is far from an estimate from 2008 of dietary fiber intake of 15.9 grams per day (King, Mainous, and Lambourne, 2012) and is likely due to not covering all possible sources of dietary fiber sources in this research.

Table 2.1. Conditional Summary Statistics for Fiber Intake Categories (Daily Grams Per Capita)

	Bread	Pasta	Tortilla	Canned Fruit	Fresh Fruit	Frozen Fruit	Canned Vegetables	Fresh Vegetables	Frozen Vegetables
Mean	1.15	0.40	0.33	0.34	0.93	0.23	1.06	0.95	0.62
Std. Dev.	1.14	0.45	0.66	0.51	1.32	0.52	1.43	0.98	0.81
Min.	0	0	0	0	0	0	0	0	0
Max.	45.53	13.89	82.74	34.89	91.98	14.44	83.90	30.70	43.39
Percent Zero Observations	10.8	40.1	70.3	43.4	29.1	87.4	19.4	11.8	28.0

Note: This table lists summary statistics conditional on purchasing in that category.
Source: Calculated by author.

Figure 2.1. Distribution of Average Daily Fiber per Capita Intake



Note: The black dashed vertical line at 4.38 represents the sample average daily fiber intake, grams per capita.
Source: Produced by author.

Table 2.2. Summary Statistics for Observed and Imputed Prices for each of the Products

	Observed Price ($/gram)		Imputed Price ($/gram)	
	Mean	Std. Dev.	Mean	Std. Dev.
Bread	0.00174	0.00080	0.00176	0.00076
Pasta	0.00139	0.00085	0.00146	0.00067
Tortilla	0.00189	0.00112	0.00207	0.00066
Canned Fruit	0.00126	0.00059	0.00129	0.00045
Fresh Fruit	0.00181	0.00154	0.00187	0.00130
Frozen Fruit	0.00291	0.00109	0.00303	0.00043
Canned Vegetable	0.00111	0.00102	0.00113	0.00091
Fresh Vegetable	0.00216	0.01150	0.00216	0.01080
Frozen Vegetable	0.00170	0.00083	0.00169	0.00070

Note: Imputed prices were calculated with an auxiliary regression that included household income, household size, location, and time variables.
Source: Calculated by authors.

For the explanatory variables, we begin by including prices for each of the nine product categories. Unit values (proxy for prices) are calculated by taking the total expenditure in a category and dividing this by the total weight (grams) purchased for that category. We observe no unit value or price for the transactions with zero quantities and hence zero expenditures (due to censored nature of these observations). Missing prices are imputed using an auxiliary regression of quantity purchased on household income, household size, location, and time variables. The variable for income controls for different levels of quality while the other variables account for price differences caused by regional differences, demographic variability and time effects. This approach is not without precedent and is standard procedure used in the price imputation literature (Capps, et al, 1994; Alviola and Capps, 2010, Kyureghian, Nayga and Capps, 2011, and Dharmasena and Capps, 2014). In table 2.2, we present summary statistics for observed prices and imputed prices for each of the product categories. For those wondering about the consistency of the observed and imputed prices, the table shows the means of the imputed prices to be close to the means of the observed prices.

Then we include a range of standard demographic characteristics to uncover the effect of various household characteristics affecting the demand for fiber intake. Table 2.3 lists summary statistics for these characteristics and some additional household characteristics not directly included in the estimation procedure. The age variable was constructed to consider only the age of the oldest head of the household. It is assumed that the oldest member is likely to have more influence on purchase decisions. The average age of the sample is 60 years.[4] This sample has 5% of respondents identifying as being of Hispanic origin. Controlling for Hispanic origin and races is important because such respondents may have different preferences over the categories based on socio-cultural factors of individuals (e.g., more likely to consume tortillas for those of Hispanic origin).

We construct an indicator for the presence of children in the households. This variable indicates if there is at least one child present in the household. This may be important as the presence for children may change the nutritional mix of food purchased. Parents may focus on purchases of healthier food when children are present in the household.

[4] This result shows that the sample constructed is not representative of the U.S. population because the sample was constructed to follow only those individuals who participated in the panel for 11 years. Further research using the Nielsen provided survey weights might make the sample more representative of the U.S. population.

Table 2.3. Summary Statistics and Household Characteristics

Variable	Mean	Std. Dev.
Real Unit Prices ($/100 grams)		
Bread	0.18	0.02
Pasta	0.15	0.02
Tortilla	0.21	0.03
Canned Fruit	0.13	0.01
Fresh Fruit	0.19	0.04
Frozen Fruit	0.30	0.02
Canned Vegetables	0.11	0.02
Fresh Vegetables	0.17	0.05
Frozen Vegetables	0.17	0.01
Demographic characteristics		
White	0.84	0.36
Black	0.09	0.29
Asian	0.03	0.16
Other	0.04	0.19
Hispanic origin (any race)	0.05	0.21
Age of oldest head of household	60.19	11.47
Economic Characteristics		
Real household income	25,963	13,277
Income below 130% of poverty line	0.07	0.26
Income between 130% and 185% of poverty line	0.16	0.16
Education		
Less than HS degree	0.01	0.12
HS degree	0.20	0.40
Some college	0.28	0.45
Bachelor's or higher degree	0.50	0.50
Family Characteristics		
At least one child present	0.13	0.33
Household Size	2.08	1.10
Place of residence		
Northeast	0.17	0.38
Midwest	0.26	0.44
South	0.35	0.48
West	0.22	0.42

Note: This table lists summary statistics for the explanatory variables. We report the mean for each characteristic and standard deviations. Categories may not sum to one due to rounding. Except for prices, family income, and household size, all characteristics are indicators.
Source: Calculated by author.

Household income is included in the set of explanatory variables. Nielsen provides only categorical income information. The income variable is constructed as the natural log of the midpoint of the categorical yearly income variable. The average for real income is $27,100. The poverty dummy variables indicate whether the respondent's household income is at or below 130% of the federal poverty level and whether the household is between 130% and 185% of the federal poverty level. These variables are calculated using household income and household size. This is a reason why the two variables do not appear in the final regression. The threshold levels here indicate eligibility for participation in public assistance programs such as the Supplemental Nutritional Assistance Program (SNAP) at 130% and below or WIC at above 130% and below 185%. By using indicator variables for eligibility, rather than indicators for actual participation, we avoid potential complications arising from non-random selection into the programs and under-reporting of participation.

The place of residence dummies use the four U.S. Census Bureau designated divisions. These are used to control for possible differences in the characteristics of the food environment. These differences may include the availability of grocery stores and other food outlets or possible geographical differences in food tastes and preferences.

Results

Table 2.4 presents the unconditional marginal effects while table 2.5 presents the conditional marginal effects. Table 2.6 presents the unconditional elasticities while table 2.7 presents the conditional elasticities. The marginal effect of an explanatory variable illustrates the influence of a change in this variable on the expected intake of fiber (conditional and unconditional) from a given product, by accounting for all impacts associated with the change in the variable. The conditional marginal effects and elasticities are the changes in the intake of dietary fiber given that the household intakes a positive amount of dietary fiber in that category. We focus on the conditional marginal effects and elasticities when discussing the results below as these explain the effects of the explanatory variables on fiber intake for those that actually made a transaction. The elasticities used in this study are interpreted as the percentage change in the intake of fiber from a given product given a 1% change in the price of that product. Due to the large sample size, significance is considered only at the 1% level (p-value 0.01). The estimation results reveal statistically significant effects of economic variables on the fiber purchase categories.

The indicator for the 2010 Dietary Guidelines seems to show mixed results. Six of the fiber categories show a negative effect while three of the categories show a positive effect. For the mean household in the sample these results imply at 5.5% decrease in fiber in the time period after the dietary guidelines were released. Fiber intake from bread shows the largest negative effect. Negative consumer perception about carbohydrates may have decreased bread consumption, which led to a decrease in the intake of fiber from bread. Fiber from fresh fruit shows the largest positive effect. Given these results, it is likely that our sample was not persuaded by the guidelines to significantly alter their diets in order to intake more fiber.

Table 2.4. Unconditional Marginal Effects of the Panel Tobit Model

	Bread	Pasta	Tortilla	Canned Fruit	Fresh Fruit	Frozen Fruit	Canned Vegetables	Fresh Vegetables	Frozen Vegetables
Guidelines 2010 Indicator	**-0.091**	**-0.006**	**0.021**	**-0.053**	**0.093**	**0.015**	**-0.048**	**-0.011**	**-0.031**
Below 130% Poverty Level	**-0.025**	**-0.011**	**-0.012**	**-0.016**	**-0.025**	**-0.015**	**-0.031**	**-0.017**	-0.008
130% to 185% Poverty Level	**-0.029**	**-0.014**	**-0.017**	-0.002	**-0.054**	**-0.010**	**-0.028**	**-0.035**	**-0.028**
Hispanic [a]	-0.020	-0.009	**0.094**	**-0.020**	0.010	0.016	0.021	0.015	0.000
Black [b]	**-0.050**	**-0.058**	**-0.130**	**-0.052**	**-0.098**	**-0.065**	**-0.124**	**-0.159**	**0.063**
Asian [b]	**-0.169**	-0.008	**-0.037**	**-0.068**	0.065	**-0.040**	**-0.162**	**-0.075**	**-0.052**
Other [b]	**-0.027**	-0.011	0.009	**-0.015**	-0.015	-0.005	-0.028	**-0.046**	-0.005
Age of oldest head of household	**0.001**	**0.001**	**-0.003**	**0.002**	**0.012**	**0.001**	**0.004**	**0.003**	0.001
Child present	**-0.212**	**-0.046**	**-0.016**	**-0.031**	**-0.114**	**-0.016**	**-0.208**	**-0.209**	**-0.086**
West [c]	**0.042**	-0.004	**0.175**	0.003	-0.006	**0.017**	**-0.125**	**-0.034**	**-0.090**
Midwest [c]	-0.001	**0.026**	-0.011	**0.030**	**0.092**	0.006	**-0.094**	**0.039**	**-0.040**
Northeast [c]	**0.066**	**0.129**	**-0.063**	0.007	0.025	-0.009	**-0.104**	**0.070**	0.021

Note: This table presents the estimated unconditional marginal effects of the panel Tobit model. Bold represents significance at the 1% level (p-value 0.01).

a Base category is non-Hispanic origin

b Base category is White

c Base category is South region

Source: Calculated by author.

Table 2.5. Conditional Marginal Effects of the Panel Tobit Model

	Bread	Pasta	Tortilla	Canned Fruit	Fresh Fruit	Frozen Fruit	Canned Vegetables	Fresh Vegetables	Frozen Vegetables
Guidelines 2010 Indicator	**-0.067**	**-0.003**	**0.008**	**-0.024**	**0.051**	**0.005**	**-0.031**	**-0.008**	**-0.016**
Below 130% Poverty Level	**-0.019**	**-0.005**	**-0.005**	**-0.007**	**-0.013**	**-0.005**	**-0.020**	**-0.011**	-0.004
130% to 185% Poverty Level	**-0.022**	**-0.006**	**-0.006**	-0.001	**-0.030**	**-0.003**	**-0.018**	**-0.024**	**-0.015**
Hispanic [a]	-0.015	-0.004	**0.034**	**-0.009**	0.006	0.005	0.014	0.010	0.000
Black [b]	**-0.037**	**-0.027**	**-0.047**	**-0.023**	**-0.053**	**-0.022**	**-0.079**	**-0.108**	**0.033**
Asian [b]	**-0.125**	-0.004	**-0.013**	**-0.030**	0.036	**-0.013**	**-0.104**	**-0.051**	**-0.027**
Other [b]	**-0.020**	-0.005	0.003	**-0.007**	-0.008	-0.002	-0.018	**-0.031**	-0.002
Age of oldest head of household	**0.001**	**0.000**	**-0.001**	**0.001**	**0.007**	**0.000**	**0.003**	**0.002**	0.000
Child present	**-0.157**	**-0.021**	**-0.006**	**-0.014**	**-0.062**	**-0.005**	**-0.133**	**-0.141**	**-0.045**
West [c]	**0.031**	**-0.002**	**0.064**	0.001	-0.003	**0.006**	**-0.080**	**-0.023**	**-0.047**
Midwest [c]	-0.001	0.012	-0.004	**0.013**	**0.050**	0.002	**-0.060**	**0.026**	**-0.021**
Northeast [c]	**0.049**	**0.059**	**-0.023**	0.003	0.014	-0.003	**-0.066**	**0.047**	0.011

Note: This table presents the estimated conditional marginal effects of the panel Tobit model. Bold represents significance at the 1% level (p-value 0.01).
a Base category is non-Hispanic origin
b Base category is White
c Base category is South region
Source: Calculated by author.

Table 2.6. Unconditional Elasticities of Demand for Fiber Generated from the Panel Tobit Model

	Bread	Pasta	Tortilla	Canned Fruit	Fresh Fruit	Frozen Fruit	Canned Vegetables	Fresh Vegetables	Frozen Vegetables
Ln Bread unit price	**-0.150**	**0.014**	**0.024**	-0.001	**0.058**	**0.026**	**0.033**	**0.013**	**0.013**
Ln Pasta unit price	**-0.026**	**-0.246**	**0.010**	**-0.005**	**0.012**	**0.014**	**-0.024**	**-0.014**	0.001
Ln Tortilla unit price	**-0.019**	**-0.007**	**-0.317**	**-0.009**	**0.015**	0.004	**-0.032**	**-0.025**	**-0.010**
Ln Canned Fruit unit price	**-0.020**	-0.004	0.002	**-0.182**	**-0.020**	-0.005	**-0.052**	**-0.021**	**-0.020**
Ln Fresh Fruit unit price	**-0.023**	**-0.009**	**-0.004**	**-0.022**	**-0.436**	**-0.006**	**-0.047**	**-0.044**	**-0.018**
Ln Frozen Fruit unit price	0.017	**0.037**	0.006	**0.040**	0.013	**-0.300**	**0.112**	**0.025**	**0.048**
Ln Canned Vegetables unit price	**-0.029**	**-0.011**	-0.004	**-0.024**	-0.007	0.002	**-0.532**	**-0.035**	**-0.008**
Ln Fresh Vegetables unit price	**-0.039**	**-0.020**	**-0.006**	**-0.027**	**-0.012**	0.001	**-0.067**	**-0.338**	**-0.028**
Ln Frozen Vegetables unit price	**-0.018**	**-0.005**	**-0.005**	**-0.013**	**0.015**	0.004	**-0.030**	**-0.032**	**-0.197**

Note: This table presents the estimated unconditional elasticities of the panel Tobit model. Bold represents significance at the 1% level (p-value 0.01).
Source: Calculated by author.

Table 2.7. Conditional Elasticities of Demand for Fiber Generated from the Panel Tobit Model

	Bread	Pasta	Tortilla	Canned Fruit	Fresh Fruit	Frozen Fruit	Canned Vegetables	Fresh Vegetables	Frozen Vegetables
Ln Bread unit price	**-0.099**	**0.017**	**0.030**	-0.001	**0.036**	**0.048**	**0.021**	**0.009**	**0.012**
Ln Pasta unit price	**-0.017**	**-0.298**	**0.012**	**-0.007**	**0.007**	**0.025**	**-0.015**	**-0.010**	0.000
Ln Tortilla unit price	**-0.012**	**-0.008**	**-0.399**	**-0.013**	**0.009**	0.007	**-0.020**	**-0.018**	**-0.009**
Ln Canned Fruit unit price	**-0.013**	-0.005	0.003	**-0.254**	**-0.012**	**-0.008**	**-0.032**	**-0.015**	**-0.017**
Ln Fresh Fruit unit price	**-0.015**	**-0.010**	**-0.005**	**-0.031**	**-0.269**	**-0.012**	**-0.030**	**-0.032**	**-0.016**
Ln Frozen Fruit unit price	0.011	**0.044**	0.007	**0.056**	0.008	**-0.551**	**0.070**	**0.018**	**0.042**
Ln Canned Vegetables unit price	**-0.019**	**-0.014**	-0.005	**-0.033**	-0.004	0.004	**-0.333**	**-0.026**	**-0.007**
Ln Fresh Vegetables unit price	**-0.026**	**-0.024**	**-0.007**	**-0.037**	**-0.007**	0.001	**-0.042**	**-0.245**	**-0.025**
Ln Frozen Vegetables unit price	**-0.012**	**-0.006**	**-0.006**	**-0.018**	**0.010**	0.008	**-0.019**	**-0.023**	**-0.173**

Note: This table presents the estimated conditional elasticities of the panel Tobit model. Bold represents significance at the 1% level (p-value 0.01).
Source: Calculated by author.

Focusing on the own-price elasticities of fiber demand shows that for all categories, fiber intake is inelastic. Our estimates of the own-price elasticities for fiber from fruit range from -0.25 for canned fruit to -0.55 for frozen fruit. Our estimates for the own-price elasticities for fiber from vegetables range from -0.17 for frozen vegetables to -0.33 for canned vegetables. As frozen vegetables are the most inelastic produce item, any subsidy will be least effective at increasing fiber intake if targeting this product. As a reminder on how to interpret these elasticities using fresh vegetables as an example a 1% increase (decrease) in the price of fresh vegetables will result in a 0.25% decrease (increase) in the intake of dietary fiber, derived from fresh vegetables.

These results are not directly comparable with those of other studies, since they focus on the fruit or vegetable as the object of quantity demanded and not the quantity of a nutrient, the dietary fiber considered in this study. It is still informative to compare with elasticity estimates for fruit and vegetables and a change in the consumption of that item will lead to a change in dietary fiber intake. Park et al. (1996) find own price elasticities of -0.34 for fruit and -0.32 for vegetables for low-income households. Dong and Lin (2009) find own-price elasticities of -0.52 for fruit and -0.69 for vegetables for low-income households.

The cross-price elasticities for the fiber intake between products are much weaker. All cross-price elasticities for bread are negative with vegetables having the strongest negative elasticity. Pasta fiber intake has positive cross-price elasticities for bread and frozen fruit and negative for all others. Tortilla fiber intake has a positive cross-price elasticity with bread and pasta fiber and negative for fresh fruit, fresh vegetables, and frozen vegetables. For those obtaining their fiber from tortillas, bread and pasta fiber intake may be gross substitutes.

Canned fruit fiber has negative cross-price elasticities for all categories except frozen fruit. It is not surprising that frozen fruit fiber may be a gross substitute for canned fruit fiber. Fiber intake from fresh fruit has positive cross-price elasticities except for canned fruit fiber and fresh vegetable fiber. Frozen fruit fiber intake has a positive cross-price elasticity with bread and negative for canned fruit and fresh fruit. These results provide little evidence that that the various forms of fruit are gross substitutes in fiber intake.

Canned vegetable fiber has negative cross-price elasticities expect for bread and frozen fruit. Fiber from fresh vegetables have negative elasticities for all except for bread and frozen fruit. The elasticities for frozen vegetables are all negative except for bread and frozen fruit. Again these results indicate that the various forms of vegetable fiber are not substitutes for each other. Targeting one form of produce will not cause people to stop consuming that form of produce.

Households below the poverty indicators do significantly differ in their fiber purchases in most categories. The largest effects are in the bread, fresh fruit, canned vegetables, and fresh vegetables. While our results find a negative effect for WIC eligibility on fiber intake in some categories, Miguel and Diansheng (2012) find that participation in the WIC program has no effect on dietary fiber intake overall.

Some further interesting results arise from the household characteristics. We find that the presence of children in the household is associated with lower per capita fiber intake across all the categories, which was not expected. This may be caused by only including an indicator for the presence of a child. Further refinement by including categories for different ages of children may reveal other effects as children requires less fiber than adults.

Also, there are interesting age effects in the results though these are small. Older individuals intake more fiber per capita from bread and fresh vegetables relative to younger households. Additionally, there appear to be regional differences in fiber intake. For example, the Northeast region has a higher intake of fiber from bread and pasta than the South region.

Table 2.8. Correlation of Panel Tobit Residuals

	Bread	Pasta	Tortilla	Canned Fruit	Fresh Fruit	Frozen Fruit	Canned Vegetable	Fresh Vegetable	Frozen Vegetable
Bread	1.00								
Pasta	0.12	1.00							
Tortilla	0.05	0.05	1.00						
Canned Fruit	0.07	0.08	0.03	1.00					
Fresh Fruit	0.09	0.07	0.03	0.04	1.00				
Frozen Fruit	0.00	0.02	0.06	0.07	0.04	1.00			
Canned Vegetable	0.17	0.19	0.07	0.24	0.10	0.04	1.00		
Fresh Vegetable	0.12	0.19	0.07	0.07	0.17	0.03	0.23	1.00	
Frozen Vegetable	0.12	0.16	0.05	0.08	0.11	0.08	0.19	0.14	1.00

Note: This table provides the correlation between the residuals of the nine panel Tobit equations. The table shows that these correlations are low and that it is likely no procedure are needed to account for a correlation among error terms.
Source: Calculated by author.

Because there is a possibility that the fiber intake categories are likely to be affected by the same set of unobservable factors, we need to check the correlation of the residuals. Table 2.8 presents the correlation of the panel Tobit residuals. By examining these residuals, we see that many are very low. This ensures that a seemingly unrelated regression (SUR) approach which considers a system of equations is not warranted (Zellner, 1962). Due to this, the equations can be efficiently estimated equation by equation. For more details on estimating a SUR Tobit model, see Huang (2001).

Table 2.9 shows the effects of a 20% price decrease due to a subsidy, which would reduce the price consumers pay, applied to four scenarios. Scenario 1, a 20% subsidy applied to all categories of fruits and vegetables, would result in an increase in the average per capita intake of fiber per day by 4.8%. Scenario 2, this subsidy applied to only canned products, would result in a 2.1% increase. Scenario 3, the subsidy applied to only fresh products, would result in a 2.7% increase. Scenario 4, the subsidy applied only to frozen produce would result in a 0.05% increase in per capita average daily fiber intake. A subsidy targeting only fresh produce would lead to a much higher increase in fiber intake relative to a subsidy targeted at only frozen produce. Targeting frozen produce will give the least return per dollar invested.

Table 2.9. Percent Change in Grams/Day Fiber Intake from a Proposed 20% Price Decrease due to a Subsidy

	Bread	Pasta	Tortilla	Canned Fruit	Fresh Fruit	Frozen Fruit	Canned Vegetables	Fresh Vegetables	Frozen Vegetables	Total Percent Change
Scenario 1	1.48	0.29	0.28	6.33	5.51	11.16	7.73	6.46	3.91	4.80
Scenario 2	0.64	0.38	0.04	5.75	0.33	0.08	7.32	0.81	0.48	2.06
Scenario 3	0.83	0.69	0.25	1.36	5.53	0.21	1.44	5.54	0.81	2.69
Scenario 4	0.01	-0.77	-0.01	-0.77	-0.35	10.87	-1.03	0.10	2.61	0.05
Baseline grams/day	0.90	0.22	0.09	0.16	0.75	0.04	0.82	0.83	0.42	

Note: Scenario 1 is a 20% subsidy applied to all fruit and vegetables. Scenario 2 is this subsidy applied to only canned fruit and vegetables. Scenario 3 is the subsidy applied to only fresh fruit and vegetables. Scenario 4 is the subsidy applied to only frozen fruit and vegetables. The baseline grams per day is the average of last 4 quarters of the per capita fiber intake in the respective category. This baseline amount is increased or decreased by the corresponding conditional own and cross price elasticities to find the percent change for each category. The total percent change is the difference from the total baseline amount of 4.23 grams per day and the amount after the subsidy is applied.
Source: Calculated by author.

Table 2.9 shows that the effects of the 20% subsidy to be disappointing in trying to meet the dietary fiber intake guideline. Therefore, it would be further informative to determine the amount of subsidy required to meet the daily guideline. Assuming the elasticities do not change over the range of the subsidy and given a target of 25 grams per day, we can determine this subsidy by solving a linear programming problem. We seek to get the daily fiber intake to 25 grams per day by finding a subsidy level to meet this level conditional on the set of conditional elasticities given in table 2.7. A subsidy of 1064% applied to all type of fruits and vegetables, *cetris paribus*, would be necessary to get 25 grams per day. In addition, a subsidy of 2755% applied to only fresh fruits and vegetables, *cetris paribus*, would meet the guideline. These results show that a subsidy alone would not be effective in getting consumers to meet the dietary fiber intake guideline.

Conclusions, Implications, and Limitations

By conducting a panel regression on nine per-capita fiber categories taken from purchases of bread, pasta, tortilla, fresh fruit, fresh vegetables and beans, frozen fruit, frozen vegetables and beans, canned fruit, and canned vegetables and beans this paper sought to uncover socioeconomic and government food policy related factors on the per capita intake of dietary fiber in the United States. Furthermore, an inquiry into a 20% subsidy on the aforementioned food categories to encourage the intake of fiber is conducted. A number of interesting findings result from the analysis. The results indicated that those living below 130% and those between 130% and 185% of the poverty level purchase less fiber per capita relative to those above these poverty levels.

For the mean household in the sample there is a 5.5% decrease in per capita dietary fiber purchased in the time period after the 2010 dietary guidelines were released compared to before the release The results indicate that the various forms of vegetable and fruit fiber are likely not substitutes for each other. A proposed 20% subsidy applied to all categories of fruits and vegetables would result in an increase in the average per capita intake of fiber per day by 4.8%. This subsidy when applied to only canned products would result in a 2.1% increase, applied to only fresh products would result in a 2.7% increase, and applied only to frozen would result in a 0.05% increase in per capita average daily fiber intake. In addition, a subsidy of 2755% applied to only fresh fruits and vegetables, *cetris paribus*, would be necessary to meet a guideline of 25 grams per day. Thus, subsidies alone would not be easily able to encourage consumers to meet the daily fiber intake guideline.

While we believe we accounted for all possible sources of bias in our modeling procedure, limitations still remain. The expected issues from self-reported data and the restriction to Nielsen households currently prevent generalization to all households.[5] Given the inherent limitation attributed to Nielsen Homescan Consumer Panel data, the focus of this paper is food purchased for consumption at home. Fiber intake away from home would not be captured by this dataset. This may not be a major problem as eating meals away from home is usually associated with less healthy eating (Lin and Guthrie, 2012; Todd, Mancino and Lin, 2010) and this might not change overall fiber totals. Also,

[5] Einay, Leibtag, and Nevo (2010) have formulated a method to help correct for possible entry errors in the dataset.

76% of the dietary fiber is provided by foods consumed at home with the remaining provided by foods consumed away from home (USDA, ERS, 2014).

This dataset does not provide time spent preparing food and only includes food purchases. The need to account for time is especially important since food prices influence food production and time allocation decisions (Aguiar and Hurst, 2007; Senia, Jensen, and Zhylyevskyy, 2017). One must be careful to differentiation between food that is purchased and food that is consumed.[6] The results of this study can best be interpreted as purchase amount decisions and not consumption amount decisions.[7]

[6] The data do not provide information on food that is purchased and given away or food waste.

[7] Although an attempt has been made to ensure the distinction in the paper, consumption and purchase may have been used interchangeably. The food items purchased in this paper are usually ready to eat and need little preparation time. Thus, time inputs are less likely to affect the quality of these goods.

CHAPTER III

A COMPLEX MODEL OF CONSUMER FOOD ACQUISITIONS: APPLYING ARTIFICIAL INTELLIGENCE AND DIRECTED ACYCLIC GRAPHS TO THE NATIONAL HOUSEHOLD FOOD ACQUISITION AND PURCHASE SURVEY (FOODAPS)

The interaction between the local food environment, an individual's dietary pattern, prices, health outcomes, and policy variables is complex and rarely considered in its entirety. Many studies examine the interactions of these factors but usually not together as one system. Some studies examine the link between individual and household characteristics and the local food environment (Powell, Chaloupka, and Bao, 2007). Others examine the results between individual characteristics and an individual's dietary pattern (Darmon and Drewnowski, 2008). Some even examine the link between the local food environment and an individual's dietary pattern (Moore et al., 2008) or an individual's health outcomes (Chen, Jaenicke, and Volpe, 2016). More research is needed to examine the complex interactions among all these factors, which will help shape correct policy decisions.

This research is important because the food environment and an individual's interactions within the food environment are critically important in determining an individual's dietary quality and risk of negative health outcomes such as obesity. As explained by Finkelstein, Ruhm, and Kosa (2005) there are numerous economic causes and consequences of obesity among adults and children in the United States. Obesity is a major risk factor for diabetes, cardiovascular disease, cancer, sleep apnea, nonalcoholic fatty liver disease, osteoarthritis, and other problems (Ahima and Lazar, 2013). According to Dharmasena and Capps (2012), two-thirds of adults in the United States are either overweight or obese. Ogden *et al*., (2014) shows that childhood obesity has more than doubled in children and quadrupled in adolescents in the past 30 years in the United States. This rate has slowed, as there has been no significant changes in the obesity prevalence in youth or adults between 2003-2004 and 2011-2012. On a typical day 30% of children report consuming fast food (Bowman et al., 2004). It is important to study the acquisitions of food away from home as this account for 32% of total calories consumed (Guthrie, Lin, and Frazao, 2004). Thus, policies that want to reduce the health problems associated with obesity need a full picture of the interactions among all variables.

The contribution of this research is to use the individual and household characteristics, characteristics of the local food environment, the individual's dietary pattern, prices, health outcomes, and policy variables to estimate a graphical causality structure using the National Household Food Acquisition and Purchase Survey (FoodAPS) (USDA-ERS, 2017). This is in contrast to studies that consider these variables in a fragmented approach using *a-priori* endogenous/exogenous relationships and not as a complex system that shows many possible interactions among variables considered. Then we will calculate the parameter estimates underlying the structural relationships built from the causality structure developed through a directed acyclic graph. Finally, we will make comparisons of the causality effects from the estimated directed acyclic graph and parameter estimates to current research in the field.

The estimation of a graphical causal structure is done using directed acyclic graphs (DAGs). The DAG is generated using two algorithms: Greedy Equivalence Search (GES) and Linear non-Gaussian Orientation Fixed Structure Rule Three (LOFS R3) (Chickering, 2002; Ramsey, Sanchez-Romero, and Glymour, 2014) First, the GES algorithms is run on the data to build a graphical causal structure. Then, the LOFS R3 algorithm is run on the resulting structure to orient any edges that were not oriented by the GES algorithm. The DAG is generated under assumptions made by imposing a priori knowledge on the structure.

Our main findings can be briefly summarized as follows: Asian individuals live in an environment with high concentrations of fast food and non-fast food restaurants. Hispanic individuals live in areas with a higher concentration of fast food restaurants and food stores. Obesity is less prevalent among Asian individuals. Hispanic individuals are more likely to report a fair or poor diet. Those with higher incomes are less likely to report low food security and obesity is less prevalent among individuals in high-income groups. For two products (oils and vegetables), the quantity moves the price. However, no price directly affects the product that it represents in the current time. In regards to the paths between poverty, race and food insecurity, we find a number of paths. We find that Hispanic individuals are more likely to be food insecure. There is also a direct path between the percent of poverty level and food insecurity and a path between college education and food insecurity. We find a causal chain from Black to SNAP participation to low food security. A similar casual chain also exists for Hispanic individuals to low food security via SNAP participation.

The remainder of this paper proceeds as follows. In the Literature Review section, we discuss the existing literature on diet, health, food assistance programs, and directed acyclic graphs. In the Theoretical Background section, we discuss the theory of directed acyclic graphs. In the Data section, we give a detailed description of the data and the variables. In the Estimation Procedure, we discuss the algorithms, a priori knowledge, and parameter estimates underlying the structural relationships. In the Results section, we discuss the results and compare to the current literature. Finally, in the Conclusions section we conclude and discuss limitations.

Literature Review

In the existing literature, the individual and household characteristics, characteristics of the local food environment, the individual's dietary pattern, prices, health outcomes, and policy variables are usually not considered together at the same time in a complex system. Here we provide a brief review of literature, but a more thorough comparison of current literature with our findings is presented in the results section.

Diet and Health

As diet could possibly be linked to obesity, this is a popular line of research. Beatty, Lin, and Smith (2014) look at changes in the distribution of dietary quality among adults in the United States over the period 1989–2008. They find improvements for both low-income and higher-income individuals alike with 63% of the improvement being attributed to changes in food formulation and demographics. Stewart et al. (2011) finds that low-income households spend too much money on food that is low in fruit and vegetable content and that some money should be reallocated to fruits and vegetables to satisfy government dietary guidelines.

A section of literature also argues that individual characteristics affect an individual's dietary pattern. Darmon and Drewnowski (2008) find that diets of whole grains, lean meats, fish, low-fat dairy products, and fresh vegetables and fruit are more likely to be consumed by groups of higher income and education. In contrast, the consumption of refined grains and added fats has been associated with lower income and education. Dubowitz et al. (2008) find that a higher income and education is associated with a higher level of fruit and vegetable consumption. Zagorsky and Smith (2017) find that middle class individuals eat more fast food than the poor or wealthy and that those in the poorest income quintile eat fast food much less often than those in higher quintiles.

Some literature argues that the local food environment is affected by household characteristics. Kwate (2008) argues that Black neighborhoods will have a higher share of fast food restaurants. Powell, Chaloupka, and Bao (2007) also find associations between individual characteristics and the local food environment. These authors find that Black neighborhoods have a lower availability of restaurants compared to White and Hispanic neighborhood.

Diet and Food Assistance Programs

The food consumption behavior of Supplemental Nutrition Assistance Program (SNAP) households has been extensively studied. Liu, Kasteridis, and Yen (2013) find no evidence that SNAP participation promotes consumption of less healthy food away from home. Wilde and Ranney (2000) conclude that food spending by SNAP households peaks sharply in the first three days after benefits are received. For those who conduct major grocery shopping trips only once per month (42% of all SNAP households), calorie intake drops by the fourth week of the month. Results on the relationship between SNAP participation and food security are subject to problems of selection bias and endogeneity (Gundersen and Oliveira 2001; Jensen 2002). Gundersen and Oliveira (2001) show that once one controls for adverse selection, SNAP recipients have the same probability of food insufficiency as non-recipients.

Policymakers are interested in studies about possible incentives to improve the diet and health of residents, especially low-income households. Andrews, Bhatta, and Ploeg (2013) suggest that economic incentives should be considered as an alternative to store development in food desert communities. This would include options such as allowing SNAP households to use a portion of their benefits to fund transportation to locations with more economical shopping locations. Lin, Yen, Dong, and Smallwood (2010) look at a number of ways to increase consumption of fruit, vegetables, and milk. The authors find that a 10% price subsidy would curtail consumption deficiencies by 4%–7% at an estimated cost of $734 million a year. Studies intended to promote policies need to fully understand the interactions among health, dietary patterns, individual characteristics, and the local food environment.

Directed Acyclic Graphs

Directed acyclic graphs have been used in studies ranging from mapping the integration of brain networks (Ramsey, Hanson, and Glymour, 2011; Smith et al, 2011;

Mumford and Ramsey, 2014, Ramsey, Sanchez-Romero, and Glymour, 2014) to studying the relationship money and prices (Bessler and Lee, 2002) to modeling vehicle collision with pedestrians (Davis, 2003). Few studies on consumer food demand or the food environment exist that use directed acyclic graphs. Wang and Bessler (2006) analyze U.S. meat consumption for beef, chicken, turkey, and pork. Their study focuses on price quantity endogeneity in food products. Some meat products show evidence of endogeneity while others do not. Lai and Bessler (2015) use directed acyclic graphs to examine causal relationships among retail prices, manufacturer prices, and number of packages sold for carbonated soft drink. The results show that the retail price leads to manufacturer price and quantity sold.

The most similar paper to our research is Dharmasena, Bessler, and Capps (2016). The authors use directed acyclic graphs to model the food environment in the United States. They used state level data to model the food environment in contrast to the individual level data used in this research. The results indicate that food insecurity and participation in SNAP are related but do not seem to have a direct causal link. The authors also find that poverty and SNAP participation are related through back door paths (food insecurity, unemployment, race and food taxes).

FoodAPS

The National Household Food Acquisition and Purchase Survey (FoodAPS) is still quite recent, but there have been a few studies using the data. Taylor and Villas-Boas (2016) use the FoodAPS data with a multinomial mixed logit model to estimate food store choices as a function of type and household attributes. They find that households are willing to pay between $12 and $17 per week in distance traveled for superstores, supermarkets, and fast food, while they are willing to pay significantly less for the remaining outlets. They conclude that policymakers should consider incentivizing the building of the outlets of which there is a higher willingness to pay among consumers.

Smith et al. (2016) use the FoodAPS data to examine the SNAP benefit cycle. The authors find evidence of short-run impatience and fungibility of income behaviors in SNAP participants. Wilde, Llobrera, and Ploeg (2014) use a random sample of census block groups from the FoodAPS data to examine the adequacy of the local food retail environment. The results show that census blocks with high poverty have a closer proximity to a supermarket than other blocks. Basu, Wimer, and Seligman (2016) use FoodAPS to examine the association between cost of living and nutrition among low-income individuals. The main result for is that counties with a high cost of living are associated with a worse nutrition for low-income individuals.

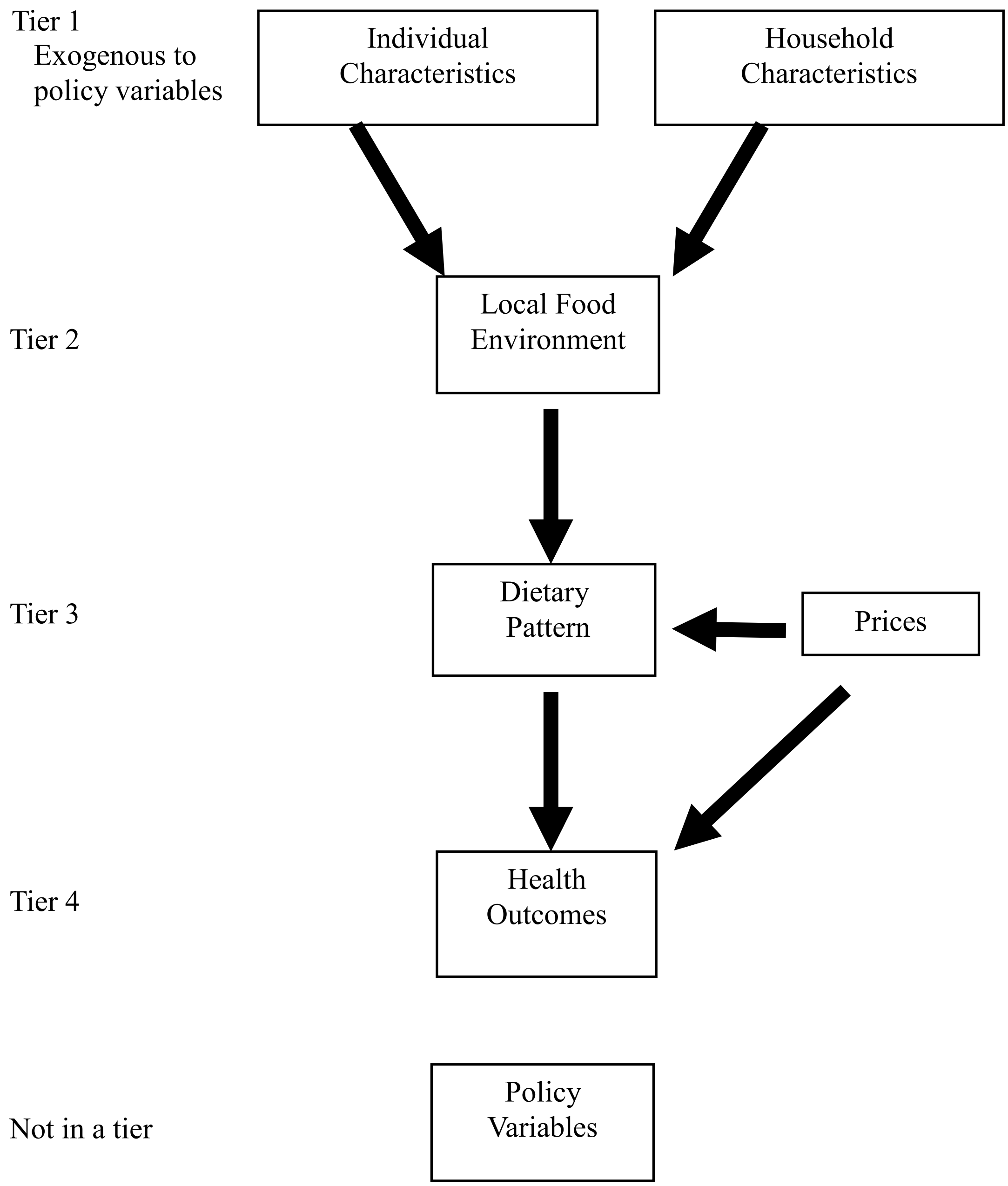
Theoretical Background

Recent literature has focused more on inferring causal relationships from observational data in the absence of controlled experiments (Pearl, 2009; Spirtes et al., 2000). These methods rely on algorithms that allow causal inferences to arise without explicitly formed hypotheses. The causal structures that arise from these algorithms can be represented in graphical form as a DAG (directed acyclic graph). The following discussion is drawn from a number of sources (Pearl, 1995; Spirtes et al., 2000; Pearl, 2009).

A graph is an ordered triple {**V**, **M**, **E**} where **V** is a non-empty set of variables (or nodes), **M** is a non-empty set of symbols attached to the ends of undirected edges, and **E** is a set of ordered pairs (edges). More simply, a graph is a diagram containing a number of nodes (which represent variables) and arrows that depict relationships among the variables. A directed graph contains only directed edges ($X \rightarrow Y$). Lines without arrows ($X - Y$) are undirected edges and are used to indicate correlations with unknown directionality. A directed acyclic graph (DAG) is a graph that does not include any cycles that start and end at the same node. For example, $X_t \rightarrow Y_t \rightarrow Z_t \rightarrow X_t$ represents a cyclic graph because it starts at the variable X and ends on the same variable X in the contemporaneous time. Figure 3.1 is an example of a directed acyclic graph. The boxes represent variables and the arrows depict the directionality of causal relationships. This graph also shows a hypothesized interaction among a set of variables extracted from the FoodAPS data, which will be modeled using DAG methods discussed later.

Nodes (variables) are sometimes referred to with the terminology of parents and children or ancestors and descendants. A parent of a node is any other node with an arrow into that node. An ancestor is any node that appears earlier than a node in a chain of nodes. A child of a node is any other node with an arrow into it from that node. A descendant is a node that occurs later in a chain.

Figure 3.1. Example Directed Acyclic Graph (DAG) with Imposed Knowledge



Note: This is an example of a DAG. The figure also demonstrates knowledge imposed during estimation. Arrows indicate the direction of causality. Not all possible paths are included for simplicity. Variable in a given tier can only affect variables in higher number tiers. The policy variables are not in a tier and can be endogenous or exogenous to any tier except that they are not allowed to cause variable in tier 1.
Source: Produced by author.

A DAG may also be referred to as a Bayesian network when its joint probability density function can be written as a product of the individual conditional density functions:

$$P(X_1, X_2, \dots, X_N) = \prod_{i=1}^{n} \mathrm{P}(X_i | pa_i) \tag{3.1}$$

where P is the probability of the variables $X_1, X_2, \dots, X_N$ and pa_i is the set of variables that precede X_i (the parents). Thus, for any set of random variables, the probability of any member of a joint distribution can be calculated from conditional probabilities using the chain rule (given an ordering of $\boldsymbol{X}$) as follows:

$$P(X_1, X_2, \dots, X_N) = P(X_1)P(X_2|X_1)P(X_3|X_2, X_1) \cdots P(X_N|X_1, X_2, \dots, X_{N-1}) \tag{3.2}$$

Pearl (1986) proposes that d-separation is a graphical version of this conditional independence.

Pearl (1995, p. 671) defines d-separation:

> Let X, Y and Z be three disjoint subsets of nodes in a directed acyclic graph G, and let p be any path between a node in X and a node in Y, where by 'path' we mean any succession of arcs, regardless of their directions. Then Z is said to block *p* if there is a node *w* on *p* satisfying one of the following two conditions: (i) w has converging arrows along p, and neither w nor any of its descendants are in Z, or, (ii) w does not have converging arrows along p, and w is in Z. Further, Z is said to d-separate X from Y, in G, written $(X \perp Y|Z)_G$, if and only if Z blocks every path from a node in X to a node in Y.

The above is a method used to read the conditional independencies of equation (3.1) directly off a graph. Geiger, Verma, and Pearl (1990) show that there is a one-to-one correspondence between the conditional independencies of equation (3.1) and the set of triples {X, Y, Z} that satisfies the *d*-separation defining in the graph G.

Further understanding of this concept can be gained by examining the three types of structures possible in a DAG: causal chains, causal forks, and colliders (inverted causal forks). A casual chain implies the variables X, Y, and Z are related as $X \rightarrow Y \rightarrow Z$. This implies a causal ordering in the variables such that X causes Y and Y causes Z (or X causes Z via Y). The correlation between X and Z will not equal zero. However, the correlation between X and Z conditional on Y will equal zero. Figure 3.1 contains the causal chain: *Local Food Environment* $\rightarrow$ *Dietary Pattern* $\rightarrow$ *Health Outcomes*. This implies that there may be a correlation between the local food environment and health outcome, but that this correlation may disappear when we condition on an individual's dietary pattern.

A casual fork implies the variables X, Y, and Z are related as $X \leftarrow Y \rightarrow Z$. This implies that Y is a common cause of X and Z. The correlation between X and Z will not equal zero. However, the correlation between X and Z conditional on Y will equal zero. Figure 3.1 does not contain any causal forks. Assume we introduce a new node to the figure such that *Local Food Environment* $\leftarrow$ *Household Characteristics* $\rightarrow$ *Nutrition Assistance Participation*. This implies that there may be a correlation between the local food environment and participation in a nutritional assistance program, but that this correlation may disappear when we condition on household characteristics.

A collider implies the variables X, Y, and Z are related as $X \rightarrow Y \leftarrow Z$. This implies that Y is caused by X and Z. The correlation between X and Z will be zero. However, the correlation between *X* and *Z* conditional on *Y* will not equal zero. Figure 3.1 contains the collider: *Local Food Environment* $\rightarrow$ *Dietary Pattern* $\leftarrow$ *Prices*. This implies that the correlation between the local food environment and prices will be zero, but if we condition on the dietary pattern then the correlation will no longer equal zero.

Three assumptions are needed to find a DAG. *Causal Sufficiency* means there are no omitted variables that can cause any of the included variables (Spirtes et al., 2000, p. 45). Next, we have the *Causal Markov condition*. This condition states that any node is independent of its non-descendants conditional on its parents (Spirtes et al., 2000, p. 53). This condition relies on *d*-separation. Finally, we have the *Causal Faithfulness condition*. This implies that any zero correlation observed between two variables is because there is no edge between these variables and is not due to cancellation of deep structural parameters (Spirtes et al., 2000, p. 56).

Data

The National Household Food Acquisition and Purchase Survey (FoodAPS) is a nationally representative panel of 4,826 U.S. households containing information about each household's food purchases and acquisitions (USDA-ERS, 2017). Details were collected about foods purchased or acquired for consumption at home and away from home and participation in nutrition assistance programs. The survey is unique in that it oversamples SNAP households and low-income households not participating in SNAP in relation to higher income households. The publicly available version of the FoodAPS dataset was used in this analysis. This means that the ERS removed all identifying characteristics and conducted a coarsening of the data. More information about their procedure can be found in the FoodAPS user guide (USDA-ERS, 2016).

FoodAPS was fielded between April 2012 and January 2013 and collected information on all food acquisitions and purchases at home and away from home by all members of the household over a seven-day period. Households had to scan barcodes, save receipts, and record other information in food journals. Information obtained from the household includes the quantities, prices, and expenditures for all at home and away from home foods and beverages purchased or acquired by all household members, eating occasions by household members. Further information was collected about household characteristics (e.g., income, program participation, food security, health status, etc.) and household access to food (e.g., location of purchase and distance to food stores and restaurants) (USDA-ERS, 2017). The USDA added information about nutrient content of purchased food and the local retail environmental based on scanned barcodes of products and household locations.

Information is available at the individual level for the 14,317 individuals who participated in FoodAPS. As obesity is a variable of interest, we restrict the sample to only those where the information is available on body mass index (BMI). This is given for 13,336 individuals in the sample. The FoodAPS survey only has BMI information for individuals 2 years old (24 months) or older. This is due to a committee on childhood obesity that concluded that for children under 2 years of age, BMI values are not helpful (Barlow, 2007). Further, we drop a few cases that are missing racial or ethnic status as this characteristic would be difficult to impute. This leaves a final sample of 13,329 individuals.

Table 3.1 gives descriptions and summary statistics for all variables used in the DAG. The first section presents individual or household characteristics. Household characteristics include the size of the household and the household's average monthly income as a percent of the poverty guideline. The average monthly income is the sum of average imputed income for each member of the household. This income is then used to find the percent of the poverty guideline given the household's characteristics. The imputation procedure and calculation as percent of the poverty guideline were given by the FoodAPS survey. The average household has 3.87 members and an income that is 236.67% of the poverty guideline. The 2012 poverty guidelines give the poverty guideline for a family of four as $23,050 (HHS, 2012). This roughly translates into an average income for the households in our sample, $54,552.

Next, a number of individual characteristics are presented in the table. All of these individual characteristics are indicator variables except for age. The average age of individuals in the sample is around 35 years. Around 53% of the sample is female. In regards to racial identification, 67% are White, 15% are Black, 4% are Asian, 1% are American Indian, and the rest identify as another race. Around 25% of the sample claim Hispanic ethnicity.

Table 3.1. Descriptions of Variables and Summary Statistics

Variable Name	Variable Description	Mean	Std. Dev.	Min	Max
Household or Individual Characteristics					
HHsize	Number of people at residence, excluding guests	3.8732	1.9658	1.0000	14.0000
PctPovGuideHH	Household average (monthly) income as sum of average imputed income per member as percent of household poverty guideline	236.6703	233.2459	0.0000	2755.5953
Female	Indicates if individual is female	0.5299	0.4991		
Age	Approximate midpoint of individual's age group	34.7921	21.4684	1.0000	85.0000
White	Individual in White racial category (base is other race)	0.6725	0.4693		
Black	Individual in Black racial category (base is other race)	0.1510	0.3581		
AmInd	Individual in American Indian or Alaskan Native racial category (base is other race)	0.0108	0.1034		
Asian	Individual in Asian or Native Hawaiian or Other Pacific Islander racial category (base is other race)	0.0420	0.2006		
Hispanic	indicates individual Hispanic (base in non-Hispanic)	0.2472	0.4314		
Employed	Individual is currently employed (base in not employed)	0.3973	0.4894		
HSgrad	Individual is high school grad (base is less HS or currently in school)	0.2380	0.4259		
SomeCollege	Individual completed some college (base is less HS or currently in school)	0.2120	0.4088		
CollegeGrad	Individual is college grad (base is less HS or currently in school)	0.1384	0.3453		
Local Food Environment Characteristics					
FF5	Number of fast food restaurants within 5 mi of household	73.0976	77.7892	0.0000	429.0000
NONFF5	Number of non-fast food restaurants within 5 mi of household	321.9569	508.4515	0.0000	3639.0000
SSSM5	Number of snap-authorized supermarkets and superstores within 5 miles of household	25.8998	43.5910	0.0000	383.0000
Rural	Indicates household lives in census rural area	0.2638	0.4407		

(continued)

Table 3.1. *(continued)*

Variable Name	Variable Description	Mean	Std. Dev.	Min	Max
Prices					
Pqa_dairy	Quality adjusted price of dairy in $/gram	0.0062	0.0231	-0.0047	1.0059
Pqa_protein	Quality adjusted price of protein in $/gram	0.0233	0.0604	-0.0204	3.1173
Pqa_grain	Quality adjusted price of grain in $/gram	0.0099	0.0137	-0.0043	0.3431
Pqa_fruit	Quality adjusted price of fruit in $/gram	0.0071	0.0315	-0.0064	1.7725
Pqa_veget	Quality adjusted price of vegetables in $/gram	0.0089	0.0110	-0.0061	0.4162
Pqa_oils	Quality adjusted price of oils in $/gram	0.0275	0.0428	-0.0045	0.8591
Quantities acquired					
Pr_dairy	Percent recommended amount of diary per day	0.4760	0.5025	0.0000	14.2052
Pr_fruit	Percent recommended amount of fruit per day	0.2957	0.3885	0.0000	4.8440
Pr_grain	Percent recommended amount of grain per day	0.8423	0.8659	0.0000	21.8428
Pr_meats	Percent recommended amount of protein per day	0.5905	0.7223	0.0000	20.5850
Pr_veges	Percent recommended amount of vegetables per day	0.3785	0.3903	0.0000	5.5386
Pr_calrs	Percent recommended amount of calories per day	0.8254	1.2207	0.0000	83.3027
Pr_oils	Percent recommended amount of oils per day	0.8494	1.3894	0.0000	39.2867
Health Measures					
LowFoodSecurity	Household level 30-day measure of food security, Indicator indicates low food security	0.2965	0.4567		
FairPoorDietStatus	Household level own assessment of health of diet, Indicator indicates rating of fair or low diet	0.2708	0.4444		
TooFewFruitVeges	Household level assessment if enough produce is consumed, Indicator indicates belief consumes too few fruits/vegetables	0.7288	0.4446		
FairPoorHealthStatus	Respondent's rating of individual's general health is fair or poor	0.1895	0.3919		
Obese	Indicator variable for obesity. For adults, obese determined by BMI ranges 30.0 and above. For children, obese is determined by ranges of BMI percentile at or above the 95th percentile.	0.3081	0.4617		
Policy Variables					
SNAPnowHH	Indicator if anyone in household is receiving SNAP benefits	0.3710	0.4831		
WICHH	Indicator if anyone in household receiving WIC benefits	0.1345	0.3412		

Note: This table contains summary statistics for all variables used in the DAGs. The variable names are the same as those in DAG figures. The variables with unreported minimums and maximums are indicator variables.
Source: Calculated by author.

A few education and employment indicators are also included in the table. Close to 40% of the sample is currently employed. This may seem low, but a number of individuals under 18 are included in this sample. These individuals be included in the unemployed category and the less than high school degree category as they are currently in school. This sample contains 41% with less than a high school degree or currently enrolled in school, 24% with a high school degree, 21% with some college completed or Associate's degree, and 14% with a college degree.

The next section of the table presents the local food environment characteristics. For the FoodAPS individuals, Todd and Scharadin (2016) find that 87% visited large grocery stores and supermarkets, and 85% visited restaurants and other eating-places at least once. Given in the table are the number of fast food restaurants, non-fast food restaurants, and SNAP-authorized supermarkets and superstores within 5 miles of the household. Within five miles of the household, there is an average of 73 fast food restaurants, 322 non-fast food restaurants, and 26 supermarkets and superstores. Around 26% of the households live in a Census designated rural area. This variable is important because these individuals will have to travel farther in order to acquire their food.

FoodAPS provides this data for a wide number of ranges, but the five-mile distance was chosen for inclusion. This should cover the range of travel dictated by much of the research in the area. A distance of 5 miles should cover around 80% of visits to sit-down restaurants and fast-food outlets with an average distance between the food establishments and homes of 2.6 miles (Liu, Han, and Cohen, 2015). Further, only considering supermarkets and supercenters for the choice of food at home store density may seem overly restrictive as it excludes convenience stores and small grocery stores. However, supermarkets and supercenters are the dominant store of choice for most U.S. households. According to Ver Ploeg et al. (2015), around 44% of households do their primary grocery shopping at supercenters, while another 45% do their primary shopping at supermarkets. Household on average travel 3.8 miles to their primary shopping store of choice.

Next in table 3.1, we present prices faced by the household. These prices are calculated for six USDA main food categories: dairy, fruit, grains, meats, vegetables, and oils. Prices are calculated at the household level as the same prices likely apply to all individuals in the household. Most items in FoodAPS are given a code to indicate the food group. For the household, the total grams acquired are summed together for each of the six categories. The total expenditure on each of the six categories are also summed for each household. Unit values (proxy for prices) are calculated by taking the total expenditure in a category and dividing this by the total weight (grams) purchased. Next, missing prices are imputed using an auxiliary regression of quantity purchased on household income, household size, and location variables. This approach is not without precedent and is standard procedure used in the price imputation literature (Capps, et al, 1994; Alviola and Capps, 2010, Kyureghian, Nayga and Capps, 2011, and Dharmasena and Capps, 2014).

The use of unit values is also likely to create bias in the form of measurement error. It is possible the aggregates are endogenous to the choice of quality. We utilize the procedure described by Cox and Wohlgenant (1986) to correct for endogeneity in prices.

For this procedure, we regress the difference between the unit price and the mean unit price for each category on a number of household demographics.

$$p_i^u - \bar{p}_i^u = \sum_j \beta_{ij} D_{ij} + v_i \tag{3.3}$$

For equation (3.3), p_i^u is the unit price for a commodity, $\bar{p}_i^u$ is the mean unit price across all households, β_{ij} is a set of coefficients to be estimated, D_{ij} is a vector of household characteristics, and v_i is the error term. The demographics used in the regression are the characteristics of each household in the sample. We include income, household size, and dummy variable indicating the region of the country. In order to get the quality-adjusted price, we used the estimated coefficients from equation (3.3) and then calculate the following,

$$\hat{p}_i = p_i^u - \sum_j \hat{\beta}_{ij} D_{ij}, \tag{3.4}$$

where $\hat{p}_i$ are the prices to be used in the estimation in place of the observed unit prices.

The prices presented in table 3.1 are the quality-adjusted prices in U.S. Dollars per gram at the household level. The quality adjusted prices range from 0.0062 $/gram for dairy to 0.0275 $/gram for oils. Fruits and vegetables has similar quality-adjusted prices at 0.0071 and 0.0089 $/gram respectively. The quality-adjusted prices are the most variable for proteins and for oils while the grains and vegetables are the least variable. Negative quality-adjusted prices are interpreted as those individuals needing to be paid in order to want to purchase that product.

Table 3.2. Recommended Calories per Day

	Males			Females		
Age	Sedentary	Mod. Active	Active	Sedentary	Mod. Active	Active
2	1,000	1,000	1,000	1,000	1,000	1,000
3	1,000	1,400	1,400	1,000	1,200	1,400
4	1,200	1,400	1,600	1,200	1,400	1,400
5	1,200	1,400	1,600	1,200	1,400	1,600
6	1,400	1,600	1,800	1,200	1,400	1,600
7	1,400	1,600	1,800	1,200	1,600	1,800
8	1,400	1,600	2,000	1,400	1,600	1,800
9	1,600	1,800	2,000	1,400	1,600	1,800
10	1,600	1,800	2,200	1,400	1,800	2,000
11	1,800	2,000	2,200	1,600	1,800	2,000
12	1,800	2,200	2,400	1,600	2,000	2,200
13	2,000	2,200	2,600	1,600	2,000	2,200
14	2,000	2,400	2,800	1,800	2,000	2,400
15	2,200	2,600	3,000	1,800	2,000	2,400
16	2,400	2,800	3,200	1,800	2,000	2,400
17	2,400	2,800	3,200	1,800	2,000	2,400
18	2,400	2,800	3,200	1,800	2,000	2,400
19-20	2,600	2,800	3,000	2,000	2,200	2,400
21-25	2,400	2,800	3,000	2,000	2,200	2,400
26-30	2,400	2,600	3,000	1,800	2,000	2,400
31-35	2,400	2,600	3,000	1,800	2,000	2,200
36-40	2,400	2,600	2,800	1,800	2,000	2,200
41-45	2,200	2,600	2,800	1,800	2,000	2,200
46-50	2,200	2,400	2,800	1,800	2,000	2,200
51-55	2,200	2,400	2,800	1,600	1,800	2,200
56-60	2,200	2,400	2,600	1,600	1,800	2,200
61-65	2,000	2,400	2,600	1,600	1,800	2,000
66-70	2,000	2,200	2,600	1,600	1,800	2,000
71-75	2,000	2,200	2,600	1,600	1,800	2,000
76 & up	2,000	2,200	2,400	1,600	1,800	2,000

Note: Estimates based on Estimated Energy Requirements (EER) equations, using reference heights (average) and reference weights (healthy) for each age-sex group. For children and adolescents, reference height and weight vary. For adults, the reference man is 5 feet 10 inches tall and weighs 154 pounds. The reference woman is 5 feet 4 inches tall and weighs 126 pounds.
Source: U.S. Department of Health and Human Services and U.S. Department of Agriculture (2015, p. 77-78).

Next in table 3.1 are the amounts acquired per day for six USDA main food categories (dairy, fruit, grains, meats, vegetables, and oils) and calories as a percent of the recommended intake per day for each individual. To construct these variables, first we needed to find the individual's calorie requirement per day. The public use data does not provide height and weight for the individuals needed to calculate a more accurate estimation of calories. Estimates of daily calorie requirements are taken from HHS and USDA (2015, p. 77-78) and presented in table 3.2. Details of the HHS and USDA assumptions are given in the notes for table 3.2. Each individual is assigned a daily calorie recommendation based on gender and age from the moderately active column.

To estimate the total amounts acquired in away from home food consumption we total the calories and contributions to the six food categories for each event. Each food away from home event lists the household members present and if any guests were present. First, we find the household share of the meal as the percent of people present at the event that are members of the household. For example, if five people are present at the meal but only four household members are among the participants, then the household's share of the meals calories and food group contributions is 80%.

Once we have the household's share of the meal, we need to find each individual's share of calories and food group contributions. The individual share is that individual's daily calorie recommendation as a percent of the total calorie recommendation for all individual in the household present at the food away from home event. For example, if an individual's daily recommendation is 2000 calories and the household's daily recommendation is 10,000 calories for those at the food away from home even, then the individual's calorie share of the event is 20% of the household share. The individual shares are used to calculate the portion of the food event's calories and contributions to the six food categories go to each individual. Then we sum across all food away from home events over the week and divide by seven to get the daily estimate of calories acquired and of the six food groups acquired.

A similar process is followed to estimate calories and food group contributions from food at home. For food at home, the household share of the food at home acquisitions is assumed to be 100%. Given the household's share of the acquisition, we need to find each individual's share of calories and food group contributions. The individual share is that individual's daily calorie recommendation as a percent of the total calorie recommendation across all individual in the household. For example, if an individual's daily recommendation is 2000 calories and the household's daily recommendation is 10,000 calories, then the individual's share of the food at home acquisitions is 20%. The individual shares are used to calculate the percent of the total food at home acquisitions that belong to the individual. This individual share is then used to calculate the calories and contributions to the six food categories that go to that individual.

Table 3.3. Recommended Dietary Pattern by Calorie Requirement

Calories	1,000	1,200	1,400	1,600	1,800	2,000	2,200	2,400	2,600	2,800	3,000	3,200
Food Group												
Vegetables (c-eq/day)	1	1.5	1.5	2	2.5	2.5	3	3	3.5	3.5	4	4
Fruits (c-eq/day)	1	1	1.5	1.5	1.5	2	2	2	2	2.5	2.5	2.5
Grains (oz-eq/day)	3	4	5	5	6	6	7	8	9	10	10	10
Dairy (c-eq/day)	2	2.5	2.5	3	3	3	3	3	3	3	3	3
Protein (oz-eq/day)	2	3	4	5	5	5.5	6	6.5	6.5	7	7	7
Oils	15	17	17	22	24	27	29	31	34	36	44	51

Note: The Healthy U.S.-Style Pattern is based on the types and proportions of foods Americans typically consume, but in nutrient-dense forms and appropriate amounts. It is designed to meet nutrient needs while not exceeding calorie requirements and while staying within limits for overconsumed dietary components.
Source: U.S. Department of Health and Human Services and U.S. Department of Agriculture (2015, p. 80-82).

To convert from the contributions to the food categories to the percent of the daily recommendations for the six food categories we use the Healthy U.S.-Style Pattern presented in table 3.3 (HHS and USDA, 2015, p. 80-82). This gives the recommended levels of consumption across a number of food groups. The totals found using the process above are divided by the daily recommendations to give the percent acquired each day as a percent of the recommended amount. For our sample, individuals on average acquire the following percent of the daily recommendations: 48% for dairy, 30% for fruit, 84% for grain, 59% for meats, 38% for vegetables, 82% for calories, and 85% for oils.

Next in table 3.1 we present health measures for the individual and household. Low food security is based on the USDA's 30-day Adult Food Security Scale. Thirty percent of households are identified as having low food security. Next is a measure of the household's own assessment of the health of their overall diet. This variable indicates that 27% of households placed themselves in the fair or poor category (two lowest of five possible responses). Then a variable indicates if the household believes they eat too few fruits and vegetables. This indicates that 73% of households believe they need to eat more fruits and vegetables. Next, we have an indicator of the individual's belief that their health is fair or poor (two lowest of five possible responses). In the sample, 19% of individual placed themselves in these categories. Then, we present and indicator if the individual is considered obese based on BMI. For adults age 18 and older, obese is determined by BMI ranges 30.0 and above. For children, obese is determined by a BMI percentile at or above the 95th percentile. For this sample, 31% of individuals are considered obese.

Finally, at the end of table 3.1 we include two policy variables. The first policy variable indicates if anyone in the household is receiving SNAP benefits and the second indicates if anyone is receiving WIC benefits. For our sample, 37% of individuals are living in households where at least one member receives SNAP benefits and 13% of individuals are living in a household where at least one member receives WIC benefits. For the U.S. in 2012, there were 46,609,000 individuals participating in SNAP (Gray, 2014, p. 12) and a total population of 313,998,379 (U.S. Census Bureau, Population Division, 2017). This would mean that 14.8% of individuals in the U.S. participated in

SNAP for the same year as our sample. Thus, the sample shows signs of oversampling of SNAP individuals.

Estimation Procedure

Finding a Graphical Structure with GES

The Greedy Equivalence Search (GES) algorithm is used to find a graphical causal structure by searching over Markov equivalence classes (Meek, 1997; Chickering, 2002). The algorithm will assign all graphs in the same equivalence classes the same score. Two graphs are equivalent (in the same equivalence class) if the DAGs are *distributionally equivalent* and *independence equivalent*. Two graphs are distributionally equivalent if under the Markov condition (the probability structure of a graph can be written with the probabilities of the variables conditionals just on the variables' parents), the graphs share the same joint probability distribution. For three variables, this reduces the search space from 25 possible DAG structures to a search over to 11 equivalence classes (Kwon and Bessler, 2011, p. 95). One example, the graphs $A \rightarrow B \rightarrow C$ and $A \leftarrow B \leftarrow C$ will have the same joint probability structure.[8] Two DAGs are independence equivalent if the independence constraints in the two DAGs are identical.[9] Further, it is assumed that the true causal model is acyclic and there are no common hidden causes existing between variables. The variable are assumed to have direct causal influence on

[8] The joint probability for $A \rightarrow B \rightarrow C$ is $P(A,B,C) = P(A)*P(B|A)*P(C|B)$. The joint probability for $A \leftarrow B \leftarrow C$ is $P(C,B,A) = P(C)*P(B|C)*P(A|B)$. Bayes' theorem is applied to this joint probability and the result is $P(C,B,A)=P(C)*P(C|B)*P(B)/P(C)*P(B|A)*P(A)/P(B)$. This simplifies to $P(C,B,A)= P(A)*P(B|A)*P(C|B)$. Thus, $P(A,B,C) = P(C,B,A)$ and they are distributionally equivalent.

[9] For the graph $A \rightarrow B \rightarrow C$, the independence constraint is $A \perp C|B$. For the graph $A \leftarrow B \leftarrow C$, the independence constraint is $A \perp C|B$. The graphs share the same independence constraint and hence are independence equivalent.

other variables in a linear manner with each variable having a Gaussian distribution. Given the above assumptions, the GES algorithm follows inclusion optimality, that the result is the most parsimonious model that contains the true model (Chickering and Meek, 2002).

The GES algorithm works by first doing a forward search and then doing a backward search. The search begins with an empty graph of all the variables. In the forward search, GES begins adding edges between nodes that increase the score. This continues until no additional edge increases the score. After this search, the algorithm compares across all equivalence classes and chooses the class with the highest score. The forward search will find the equivalence class that include the true DAG for the data. Edges are oriented according to the orientation rules described in Sprirtes et al (2000) and Meek (1995). Appendix D provides more details on these orientation rules. In the backward search, the algorithm removes edges until no single edge removal increases the score. Once no edge removal increases the score, the algorithm stops. The DAG found from the second step will be the one that best represents the data.

The scoring algorithm is important for the function of the GES algorithm. The Bayesian Information Criterion (BIC) is the most commonly used. The BIC is a measure of the marginal likelihood of the data given the graph structure and is defined as,

$$BIC = 2lnP\left(D\middle|\hat{\theta}, G\right) - c \cdot k \cdot \ln(n) \tag{3.5}$$

Where *P* represents the probability, *D* is the data, $\hat{\theta}$ is the maximum likelihood estimate, *G* represents the structure of the DAG, *c* is a penalty parameter, *k* is the number of parameters, and *n* is the number of observations. The BIC is used to balance between fit and parsimony. The penalty parameter can be used to speed up searches and reduce the number of false positive results by producing a sparser graph. Some authors suggest a discount penalty that increases as the number of edges in the true graph increases (Ramsey, 2010).

Further Orienting Edges with LOFS

Even though the GES algorithm is highly effective at orienting edges, after running the algorithm some edges may still be un-oriented (e.g., $A - B$). The LOFS (Linear non-Gaussian Orientation Fixed Structure) algorithms can be run on the GES results to orient any un-oriented edges. The LOFS works by considering higher moments of the data. A brief outline is presented here and more detail is available in other sources (Ramsey, Hanson, and Glymour 2011; Mumford and Ramsey, 2014). Three types of LOFS algorithms will be discussed: Rule 1 (*R1*), Rule 2 (*R2*), and Rule 3 (*R3*). Ramsey, Sanchez-Romero, and Glymour (2014) discuss these three rules and other alternatives. The algorithms differ from the GES in that they rely on assumptions of non-normality of the variables. Orientations are made to maximize the non-normality of variables. A scoring method commonly used for these algorithms is the Anderson-Darling statistic for normality (Anderson and Darling, 1952).[10]

The *R1* algorithm works by adding a single directed edge, testing the residuals of all possible models from adding that edge, and then selecting the model with the most

[10] The Anderson-Darling test statistic for the null hypothesis that the data follow a normal distribution is $A^2 = -n - S$. Given the ordered data Y_i, a number of observation n, and the normal cumulative distribution function Φ, then

$$S = \sum_{i=1}^{n} \frac{2i-1}{n} [ln(\Phi(Y_i)) + ln(1 - \Phi(Y_{n+1-i}))].$$

non-normal residual. For example, consider two variables A and B that are adjacent in a graph. The variable A is regressed on the empty set and on variable B. The Anderson-Darling statistic is calculated for the residuals in both regressions. If the regression of A on B has a higher Anderson-Darling statistic, then B must be a parent of A (i.e., $B \rightarrow A$) (Ramsey, Sanchez-Romero, and Glymour, 2014).

For the *R2* algorithm, consider an undirected edge between variables A and B with P_A being the candidate parents of A excluding B and P_B being the candidate parents of B excluding A. First *R2* checks if the Anderson-Darling statistic of A conditional on B and P_A is greater than then Anderson-Darling statistic of B conditional on A and P_B. Then *R2* checks if the Anderson-Darling statistic of B conditional on P_B is greater than then Anderson-Darling statistic of A conditional on P_A. If both conditions are satisfied then $A \rightarrow B$. If both conditions are reversed then $B \rightarrow A$ (Ramsey, Hanson, and Glymour 2011).

The *R3* algorithm checks whether the Anderson-Darling statistic of residuals from the regression of A on B plus the Anderson-Darling statistic of variable B is greater than the Anderson-Darling statistic of residuals from the regression of B on A plus the Anderson-Darling statistic of variable A. If this is true, then *R3* orients the edge as $B \rightarrow A$. If the reverse relationship were true, then the edge would be oriented as $A \rightarrow B$ (Ramsey, Sanchez-Romero, and Glymour, 2014).

Imposing a Priori Knowledge

In addition to the assumption inherent to the algorithms described earlier, a number of constraints can be further imposed on the search to speed estimation and produce a sparser graph. This is also important because sometimes the data cannot distinguish between two graphs. For example, the graphs $A \rightarrow B \rightarrow C$ and $A \leftarrow B \leftarrow C$ will have the same joint probability structure. The data cannot distinguish between the two graphs. We need to impose a priori information about the direction that the arrows flow. This background knowledge can be imposed in three forms.

First, edges between two variables can be forbidden. This means that the algorithm will not be allowed to connect two variables in any direction no matter what the data may imply. Second, edges between two variable can be required. This means that the algorithm will be required to connect two variables in some direction no matter what the data may imply. Third, temporal tiers may be imposed. Edges from a later tier are forbidden in earlier knowledge tiers. These knowledge tiers provide an ordering to the variables and are helpful if we know one variable or group of variables must precede another variable or group of variables.

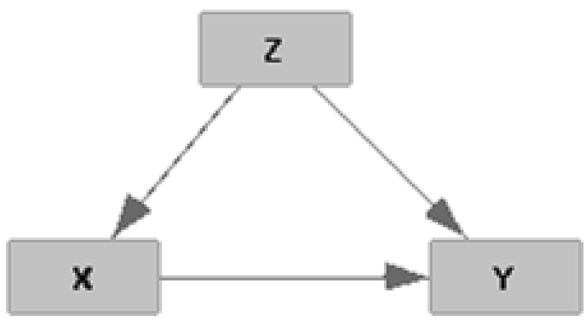
For our data, we use four knowledge tiers and some forbidden edges. Figure 3.1 provides a depiction of our four knowledge tiers. Descriptions of the variables in each category can be found in table 3.1. In the first tier, we place individual and household characteristics. These are the only variables that are required to be fully exogenous in our model. For this tier we also forbid edge within the tier. We are more interested in how these variables affect other tiers rather than the interactions among these characteristics. In the second tier, we place characteristics of the local food environment. In the third tier, we place the individual's dietary pattern and prices. These are placed in the same tier because other authors have found evidence that prices and quantities are predetermined using DAGs (Wang and Bessler, 2006). In the fourth tier, we place the health outcomes of the individual. The policy variables used in this study are not placed in any tier and are allowed to be endogenous or exogenous as determined by the algorithm. However, the policy variables are not allowed to cause the characteristics in tier 1 (forbidden edges into tier 1 variables). For example, SNAP participation is restricted not to cause an individual's characteristics (e.g., race, gender, or age).

Estimating Structural Models from Graphical Structures

After finding a DAG, one may find a corresponding structural relationship among variables that represents the graph. A short overview of estimating structural relationships is presented. A more in depth discussion on estimating these relationships can be found in Bollen (1989). The structural relationships consist of the DAG found from the algorithmic search (the variables and directed edges) and new nodes representing each error term. All edges in the graph must be directed in order to estimate the structural relationships. The causal structure of a structural relationship is indicated using directed edges from the DAG. For example, the directed edge in $A \rightarrow B$ indicates that A is the right hand side variable and B is the left hand side variable (i.e., $B = \beta A + \epsilon$). Bi-directed edges, such as $A \leftrightarrow B$, represent that the error terms between variable A and B are correlated. When constructing the structural relationship from a DAG, we assume that it is linear with Gaussian errors. A multiple linear regression is used to estimate coefficients and residual variances. The number of partial effects to be estimated is equal to the number of edges in the graph.

When estimating the parameters, two criteria are important for consideration to identify parameters (Pearl, 2009). The first to consider is the back door criteria. Suppose we are interested in the association between variables *X* and *Y* and have the graph of figure 3.2. The variables *Z* satisfy the back door criteria if: (1) no variables in *Z* are descendants of *X* and (2) *Z* blocks every path between *X* and *Y* that contains and arrow into *X*. In order to block the back door path, run a regression of *Y* on *X* and *Z* (i.e., $Y = \beta_0 + \beta_1 X + \beta_2 Z + \epsilon$). The conditioning on *Z* will block the back door and provide and unbiased and consistent estimate of $\delta Y / \delta X$.

Figure 3.2. Representation of Back Door Criteria

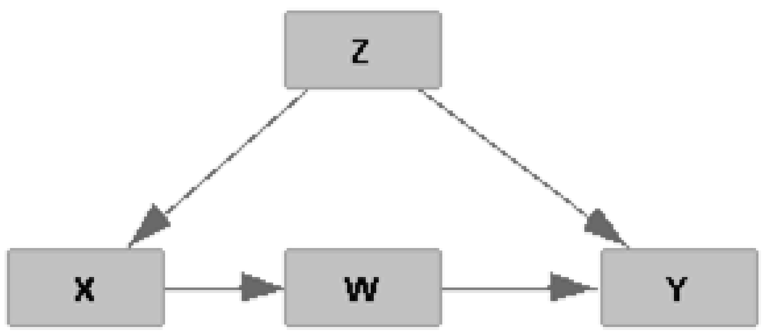


Source: Produced by author.

The next criteria for consideration is the front door criteria. Suppose we are interested in the association between variable *X* and *Y* and have the graph shown in figure 3.3. The set of variables *W* meet the front door criteria if: (1) *W* intercepts all paths directed from *X* to *Y*, (2) there are no unblocked back door paths from *X* to *W*, and

(3) all back door paths from W to Y are blocked by X. The method to block the front door path works in two steps. First, regress Y on W and X (i.e., $Y = \beta_0 + \beta_1 W + \beta_2 X + \epsilon$) to get an estimate of $\delta Y/\delta W$. Second, regress W on X (i.e., $W = \beta_0 + \beta_1 X + \epsilon$) to get an estimate of $\delta W/\delta X$. Now, and unbiased and consistent estimate of $\delta Y/\delta X$ can be found by multiplying $\delta Y/\delta W$ by $\delta W/\delta X$.

Figure 3.3. Representation of Front Door Criteria



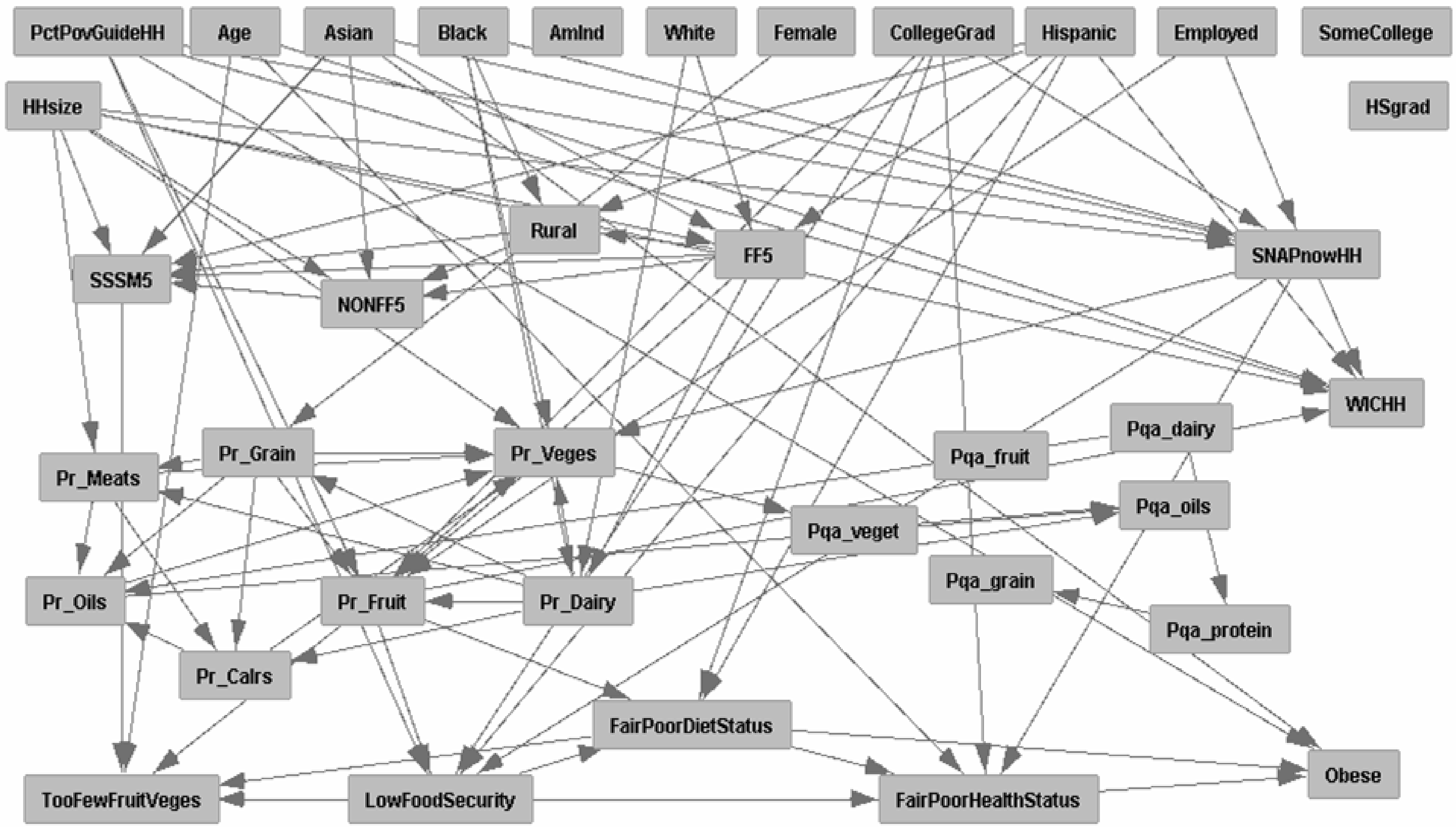
Source: Produced by author.

Results

The TETRAD V software developed by Glymour et al. (2016) is used to estimate the DAG and estimate the parameters for the structural relationships. The variables in table 3.1 are used to estimate the structure of the graph for individual and household characteristics, characteristics of the local food environment, the individual's dietary pattern, prices, health outcomes, and policy variables given the knowledge discussed earlier. After running the GES algorithm, two edges remain undirected among the quality-adjusted price variables: *Pqa_dairy* – *Pqa_protein* and *Pqa_grain* – *Pqa_protein*. The R3 LOFS algorithm is run on the graphical structure given by GES and orients the edges as *Pqa_protein* ← *Pqa_dairy* and *Pqa_protein* → *Pqa_grain*. The final graphical structure is given in figure 3.4. This figure shows the direction of causality among the variables. The partial values along with the direction for each edge can be found in table 3.4.

The DAG in Figure 3.4 shows how complicated the relationships are among the variables. It can be helpful to examine the *Markov blankets* of some variables. The Markov blanket of a node is the set of its parents, its children, and any parents of its children. This will render the variable conditionally independent from the rest of the graph. In essence, the Markov blanket of the node is the most important knowledge in predicting the behavior of a node. Figure 3.5 provides the Markov blanket for SNAP participation. Figure 3.6 provides the Markov blanket for WIC participation. Figure 3.7 provides the Markov blanket for obesity. Each of the figures also includes the partial effects from table 3.4 for quick reference. SNAP participation includes the most variables in its Markov blanket while obesity includes the least. WIC participation and SNAP participation appear in each other's Markov blanket indicating a strong dependence between the two.

Figure 3.4. Directed Acyclic Graph Associated with the Variables of Interest after Running GES Algorithm and R3 Algorithm



Note: The graph contains 88 edges between the 37 nodes. The max number of edges into a node is eight and the max number of edges out of a node is seven. The maximum total number of edges into and out of a node is 11.
Source: Produced by author.

Table 3.4. Parameter Estimates (Partial Values) for each Edge and Associated Significance

Edge		Partial			
From	To	Value	Std. Error	t-stat	p-value
Age	FairPoorHealthStatus	0.0042	0.0001	28.1328	0.0000
Age	TooFewFruitVeges	-0.0014	0.0002	-8.1234	0.0000
Age	WICHH	-0.0014	0.0001	-9.7968	0.0000
Asian	FF5	91.3143	3.1597	28.8997	0.0000
Asian	NONFF5	234.753	11.0775	21.192	0.0000
Asian	Obese	-0.1704	0.0195	-8.757	0.0000
Asian	SNAPnowHH	-0.1233	0.0192	-6.4236	0.0000
Asian	SSSM5	-8.131	1.0382	-7.8319	0.0000
Black	Pr_Dairy	-0.0992	0.0153	-6.4799	0.0000
Black	Pr_Veges	-0.0581	0.008	-7.2804	0.0000
Black	Rural	-0.1449	0.0094	-15.472	0.0000
Black	SNAPnowHH	0.1481	0.0106	13.9842	0.0000
CollegeGrad	FairPoorDietStatus	-0.0926	0.0111	-8.3485	0.0000
CollegeGrad	FairPoorHealthStatus	-0.0587	0.0094	-6.2261	0.0000
CollegeGrad	LowFoodSecurity	-0.0752	0.0115	-6.5408	0.0000
CollegeGrad	Pr_Fruit	0.0837	0.0094	8.9368	0.0000
CollegeGrad	SNAPnowHH	-0.0788	0.0119	-6.6141	0.0000
Employed	Pr_Fruit	-0.0566	0.0064	-8.7944	0.0000
Employed	SNAPnowHH	-0.0968	0.008	-12.066	0.0000
FairPoorDietStatus	FairPoorHealthStatus	0.259	0.0072	36.0712	0.0000
FairPoorDietStatus	Obese	0.0621	0.0093	6.6774	0.0000
FairPoorDietStatus	TooFewFruitVeges	0.1197	0.0087	13.7721	0.0000
FairPoorHealthStatus	Obese	0.1986	0.0105	18.8831	0.0000
Female	Pr_Grain	-0.1713	0.0135	-12.708	0.0000
FF5	NONFF5	5.8746	0.0329	178.792	0.0000
FF5	Pr_Dairy	-0.0005	0.0001	-8.7007	0.0000
FF5	Pr_Fruit	0.0004	0.0000	9.6697	0.0000
FF5	Rural	-0.0026	0.0000	-56.883	0.0000
FF5	SSSM5	-0.1524	0.0056	-27.288	0.0000
HHsize	FF5	2.3362	0.3138	7.4449	0.0000
HHsize	NONFF5	-12.769	1.1098	-11.505	0.0000
HHsize	Pr_Meats	-0.0387	0.0028	-13.696	0.0000
HHsize	Pr_Veges	-0.0111	0.0015	-7.5314	0.0000
HHsize	SNAPnowHH	0.0283	0.002	14.3775	0.0000
HHsize	SSSM5	0.7353	0.1029	7.144	0.0000
HHsize	WICHH	0.0399	0.0016	25.3534	0.0000
Hispanic	FairPoorDietStatus	0.0734	0.0088	8.3868	0.0000

(continued)

Table 3.4. *(continued)*

Edge		Partial			
From	To	Value	Std. Error	t-stat	p-value
Hispanic	FF5	61.2195	1.4495	42.2342	0.0000
Hispanic	LowFoodSecurity	0.0625	0.0087	7.1603	0.0000
Hispanic	Rural	-0.1241	0.0083	-14.992	0.0000
Hispanic	SSSM5	7.6242	0.5052	15.0926	0.0000
Hispanic	WICHH	0.0428	0.0066	6.4883	0.0000
LowFoodSecurity	FairPoorDietStatus	0.1812	0.0083	21.7309	0.0000
LowFoodSecurity	FairPoorHealthStatus	0.0778	0.0072	10.7815	0.0000
LowFoodSecurity	TooFewFruitVeges	0.0765	0.0085	8.9979	0.0000
NONFF5	SSSM5	0.0894	0.0008	113.942	0.0000
PctPovGuideHH	LowFoodSecurity	-0.0004	0.0000	-21.077	0.0000
PctPovGuideHH	Obese	-0.0001	0.0000	-7.061	0.0000
PctPovGuideHH	Pr_Fruit	0.0001	0.0000	7.3627	0.0000
PctPovGuideHH	SNAPnowHH	-0.0006	0.0000	-36.001	0.0000
PctPovGuideHH	WICHH	-0.0001	0.0000	-8.2187	0.0000
Pqa_dairy	Pqa_protein	1.7695	0.0167	106.008	0.0000
Pqa_dairy	Pr_Oils	3.1553	0.4354	7.2469	0.0000
Pqa_protein	Pqa_grain	0.0257	0.002	13.1601	0.0000
Pqa_veget	Pqa_oils	0.8572	0.0323	26.579	0.0000
Pr_Calrs	Pr_Oils	0.2089	0.0095	21.9885	0.0000
Pr_Calrs	Pr_Veges	0.0191	0.0027	7.0174	0.0000
Pr_Dairy	Pqa_oils	0.0084	0.0007	11.4947	0.0000
Pr_Dairy	Pr_Calrs	0.2925	0.0204	14.3622	0.0000
Pr_Dairy	Pr_Fruit	0.1469	0.0068	21.655	0.0000
Pr_Dairy	Pr_Grain	0.7551	0.0134	56.3933	0.0000
Pr_Dairy	Pr_Meats	0.2324	0.0122	18.9823	0.0000
Pr_Dairy	Pr_Veges	0.0619	0.0064	9.7174	0.0000
Pr_Fruit	FairPoorDietStatus	-0.0763	0.0097	-7.8721	0.0000
Pr_Fruit	LowFoodSecurity	-0.0669	0.0096	-6.9373	0.0000
Pr_Fruit	TooFewFruitVeges	-0.0909	0.0098	-9.311	0.0000
Pr_Fruit	WICHH	0.0493	0.0071	6.9252	0.0000
Pr_Grain	Pr_Calrs	0.4879	0.0124	39.401	0.0000
Pr_Grain	Pr_Fruit	0.062	0.0042	14.7555	0.0000
Pr_Grain	Pr_Meats	0.2957	0.0071	41.6817	0.0000
Pr_Grain	Pr_Oils	0.5483	0.0139	39.5445	0.0000
Pr_Grain	Pr_Veges	0.1088	0.0042	25.6402	0.0000
Pr_Meats	Pr_Calrs	0.2825	0.0141	19.9916	0.0000
Pr_Meats	Pr_Oils	0.3012	0.0157	19.2007	0.0000
Pr_Meats	Pr_Veges	0.1266	0.0045	28.1305	0.0000

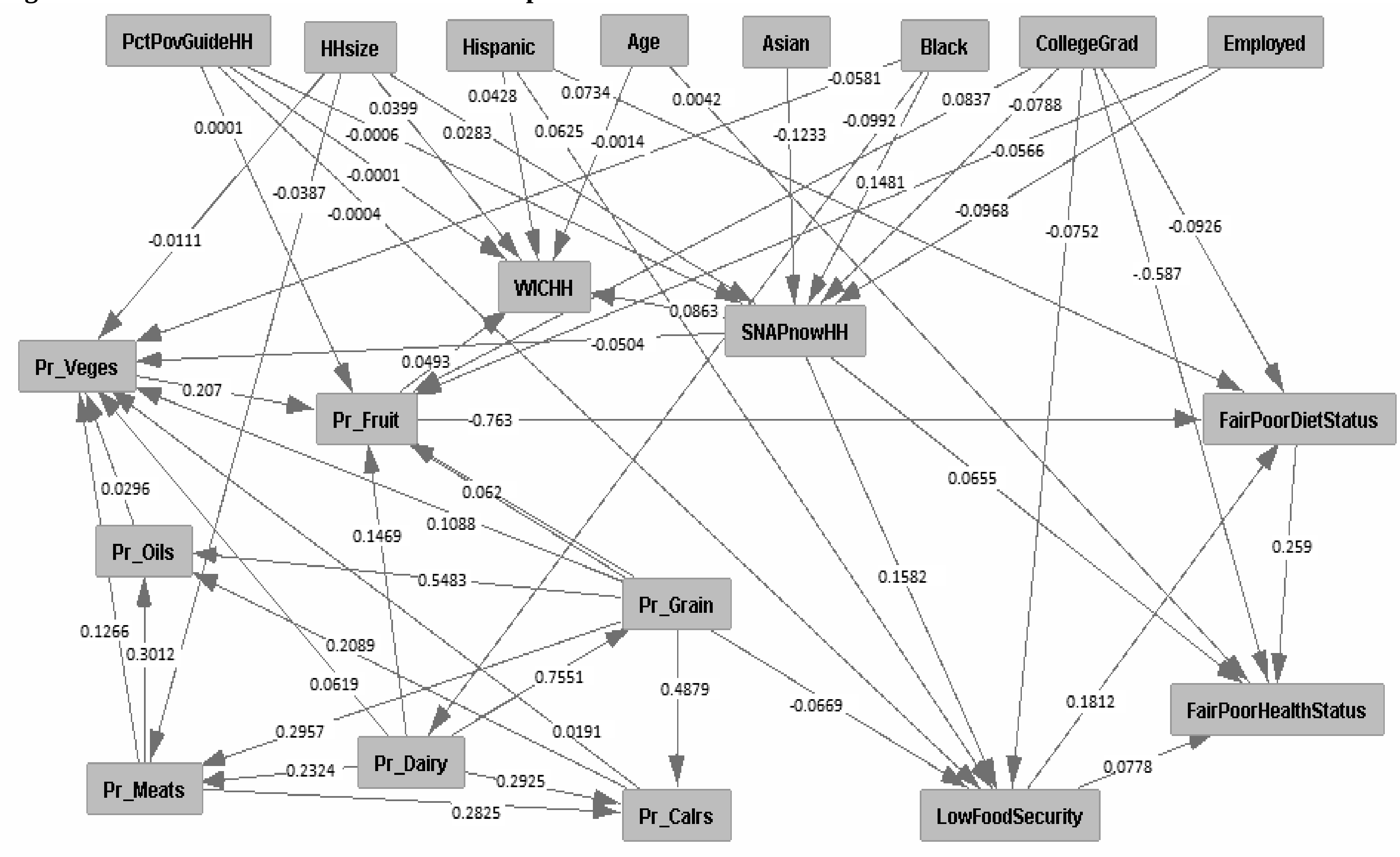
(continued)

Table 3.4. ***(continued)***

Edge		Partial			
From	To	Value	Std. Error	t-stat	p-value
Pr_Oils	Pqa_oils	0.0051	0.0003	19.3807	0.0000
Pr_Oils	Pr_Veges	0.0296	0.0024	12.2563	0.0000
Pr_Veges	Pqa_veget	-0.0019	0.0002	-7.8099	0.0000
Pr_Veges	Pr_Fruit	0.207	0.0089	23.2745	0.0000
Rural	NONFF5	123.179	5.6173	21.9284	0.0000
Rural	SSSM5	-7.4742	0.5199	-14.376	0.0000
SNAPnowHH	FairPoorHealthStatus	0.0655	0.0068	9.6373	0.0000
SNAPnowHH	LowFoodSecurity	0.1582	0.0084	18.9293	0.0000
SNAPnowHH	Pr_Veges	-0.0504	0.006	-8.4163	0.0000
SNAPnowHH	WICHH	0.0863	0.0062	13.8887	0.0000
SSSM5	TooFewFruitVeges	-0.0006	0.0001	-7.4703	0.0000
White	FF5	-22.468	1.353	-16.606	0.0000
White	Pr_Dairy	0.0985	0.0121	8.1646	0.0000

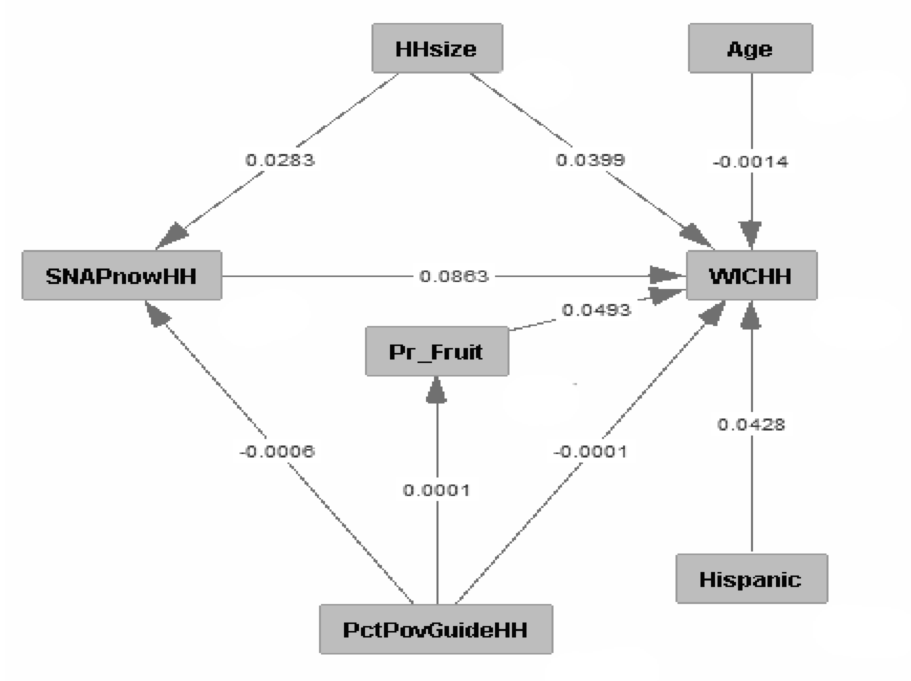
Source: Calculated by author.

Figure 3.5. Markov Blanket for SNAP Participation with Partial Effects



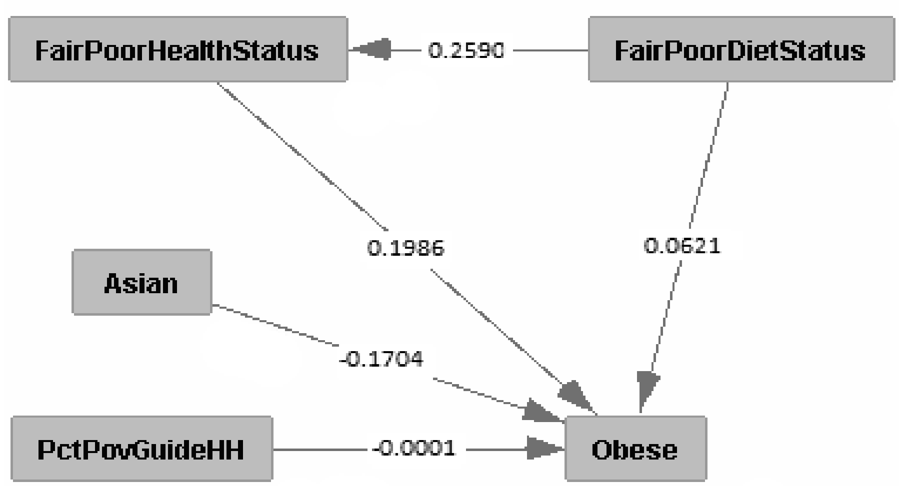
Source: Produced by author.

Figure 3.6. Markov Blanket for WIC Participation with Partial Effects



Source: Produced by author.

Figure 3.7. Markov Blanket for Obesity with Partial Effects



Source: Produced by author.

Comparisons to Other Research

Link between Individual/Household Characteristics and the Local Food Environment

Kwate (2008) argues that those classified as Black neighborhoods will have a higher share of fast food restaurants. In contrast, Powell, Chaloupka, and Bao (2007) find that predominantly Black neighborhoods have fewer full-service and fast food restaurants. The authors also find that there are significantly fewer restaurants available in predominantly Hispanic neighborhoods and that middle-income neighborhood have more restaurants than low and high-income neighborhoods.

Figure 3.4 does not show any direct path between those who are classified as Black individuals and restaurant density. The figure does show a path between Black and rural. Table 3.4 indicates that Black individuals are less likely to live in a rural area. Black does cause non-fast food restaurant density and store density through two causal chains (*Black* $\rightarrow$ *rural* $\rightarrow$ *non-fast food restaurant density* and *Black* $\rightarrow$ *rural* $\rightarrow$ *supermarket and superstore density*). Hispanic has the same two causal chains in addition to two direct connections to fast food density and store density. Including the rural variable along the non-fast restaurant and store density will block the path from the Black or Hispanic indicators to the restaurant density indicators. Asian individuals live in an environment with high concentrations of fast food and non-fast food restaurants. White individuals live in areas with a lower density of fast food restaurants. The graph also indicates a connection between household size and restaurant density and store density. Table 3.4 indicates that a larger household is more likely to live in an area with a larger density of fast food restaurants, lower density of non-fast food restaurants, and a larger selection of food stores. No paths are found between income and the local food environment variables.

Link between Individual/Household Characteristics and the Individual's Dietary Pattern

Darmon and Drewnowski (2008) find that diets of whole grains, lean meats, fish, low-fat dairy products, and fresh vegetables and fruit are more likely to be consumed by groups of higher income and education. In contrast, the consumption of refined grains and added fats has been associated with lower income and education. Little evidence indicates that income or education affects total energy intakes. Dubowitz et al. (2008) find that a higher income and education is associated with a higher level of fruit and vegetable consumption.

Figure 3.4 does show a number of paths from the individual and household characteristics to the dietary pattern variables. Table 3.4 shows a small positive effect between those with higher education and income with fruit consumption. Black individuals are found to consume less vegetables and dairy while White individuals consume more dairy. Being female leads to a lower consumption of grains. College graduates consume more fruit while those that are employed consume less fruit. A larger household size is means that individuals will eat less meat and vegetables. Our DAG does not have any other direct connections between the individual characteristics and the dietary pattern, but there are a few causal chains through the food environment variables. For example, a causal chain runs from White to fast food restaurant density to fruit consumption. The link between white and fast food density is strong and negative, and the link between fast food density and fruit consumption is small and positive. This would mean that white individuals living in areas with less fast food restaurants will consume a smaller amount of the recommended daily amount of fruit.

Link between Individual Household Characteristics and Health Outcomes

This is a more interesting area of research as some dimensions of socioeconomic status cause health, some are caused by health, and some are mutually determined with health (Cutler. Lleras-Muney, and Vogl, 2011). Smith (2007) generally finds that individuals with lower socioeconomic status have much worse health outcomes and that the primary culprit appears to be education and not an individual's financial resources. Williams et al. (2010) find that race, socioeconomic status, and gender all matter for health separately and in combination and that that racial disparities in health persist at every level of socioeconomic status.

Figure 3.4 shows a number of paths between individual characteristics and health outcomes. Table 3.4 shows that the age variable has a positive link with fair or poor health status and a negative link with too few fruits and vegetables consumed. Asian individuals are less likely to be obese and this effect is relatively strong. College graduates are less likely to report a fair or poor diet, less likely to report a fair or poor health status, and less likely to report low food security. Hispanic individuals are more likely to report a fair or poor diet and more likely to report low food security. Those with higher incomes are less likely to report low food security and less likely to be obese. The graph also indicates causal chains thorough other groups of variables. For example, from individual characteristics through dietary patterns to health outcomes (e.g., *college graduate* → *percent of recommended fruit per day* → *low food security*). A causal chain also exists from individual characteristics through dietary pattern to health outcomes (*Black* $\rightarrow$ *percent of recommended vegetables per day* $\rightarrow$ *percent of recommended fruit per day* $\rightarrow$ *low food security*). Thus, research may need to be careful conditioning on variables that might block the paths from individual characteristics to health outcomes.

Link between the Local Food Environment and the Individual's Dietary Pattern

Moore et al. (2008) finds that individuals with no supermarkets near their home are much less likely to have a healthy diet. Those with a low store density are less likely to have a healthy diet than those with a higher density. Morland, Wing, and Roux (2002) find that individuals that live in areas with a higher presence of supermarkets consume more fruits and vegetables. Timperio et al. (2008) find that for children, a higher density of fast food outlets is associates with a lower likelihood of consuming the recommended amount of fruit.

Figure 3.4 shows a few edges that connect the local food environment to the dietary pattern. The parameter estimates from table 3.4 indicate that a higher density of fast food outlets leads to consuming less dairy and more fruit, though the effects are small. In addition, a higher concentration of food stores leads to individuals being less likely to believe that their diet does not contain enough fruits and vegetables, but this effect does not show up directly in the dietary pattern. The rural indicator does not have any paths to any dietary pattern variables. That we find few edges between the local food environment and the dietary pattern variables indicates the effect of the food environment on diet is limited. This certainly is in contrast to literature that finds strong associations between the two.

Link between the Local Food Environment and Health Outcomes

Chen, Jaenicke, and Volpe (2016) find that food environment factors are associated with obesity status even after controlling for individual and household factors. Garasky, Morton, and Greder (2006) finds that the local food environment has a large impact on household food insecurity. Mehta and Chang (2008) find that a high fast food restaurant density is associated with a higher BMI and a higher full service restaurant density is associated with a lower BMI. Reitzel et al. (2014) find that a high density of fast food restaurants is positively associated with BMI but only for individuals with lower incomes.

Figure 3.4 shows only one direct path from the local food environment to health outcomes. A higher concentration of food stores leads to individuals being less likely to believe that their diet contains too few fruits and vegetables. Tale 3.4 shows that this effect is very small. There are no direct paths from the density of fast food and non-fast food restaurants to the obesity or food insecurity measure. Thus for this sample, the local food environment does not appear to have much influence on health outcomes directly. There is a causal chain from local food environment to health outcomes via the individual's dietary pattern (e.g., *fast food restaurant density* → *percent of recommended fruit per day* → *low food security*).

Link between the Individual's Dietary Pattern and Health Outcomes

Bradlee et al. (2009) find that the intakes of dairy, grains and total fruits and vegetables are inversely associated with obesity among adolescents. Wolongevicz et al. (2010) find that women with a lower diet quality are more likely to become obese than those with a higher quality diet. Wosje et al. (2010) find that a diet high in dark green and deep-yellow vegetables was associated with a lower fat mass for children.

Figure 3.4 shows that the DAG makes a connection between fruit consumption and dietary outcomes. As seen in table 3.4, individuals that consume a higher amount of fruit are less likely to report a fair or poor diet, less likely to report that they are food insecurity, and less likely to report that they consume too few fruits and vegetables. That only the fruit consumption seems to affect the health variables means that fruit consumption as a percent of the daily recommended amount may be a good proxy for food insecurity. Many of the edges stay within the dietary pattern group and do not connect to any of the health outcomes.

Link between Prices and Dietary Patterns or Health Outcomes

Wang and Bessler (2006) use DAGs to find that for some meat products, prices and quantities purchased are contemporaneous. Beydoun et al. (2011) find that a higher price index of fruits and vegetables is associated with a higher BMI. Powell and Han (2011) find that fast food and food at home prices are not associated with any broad food consumption categories.

Our price variables have mixed results in the DAG in figure 3.4. Individual and household characteristics, local food environment, and the policy variables do not have any effect on the prices. This is may be due to the procedure used to correct for quality described earlier. Only one price affects any part of the dietary pattern; the price of diary positively leads to the consumption of oils. For two items, the item's quantity causes the item's price (*percent of recommended oils per day* $\rightarrow$ *quality adjusted price of oils* and *percent of recommended vegetables per day → quality adjusted price of vegetables*). For these two items, the quantity is determined before the prices. The quantity of dairy has a link to the price of oils (*percent of recommended dairy per day → quality adjusted price of oils*). However, no price directly affects the product that it represents

Links with SNAP Participation, WIC Participation, Food Insecurity, and Obesity

Finally, we concentrate our discussion of results for some popular variables of policy interest, such as SNAP participation, WIC participation, food insecurity, and obesity to other research. We compare results from other authors to our figures 3.5, 3.6, and 3.7 for direct effects and we refer back to figure 3.4 for any causal chains. Dharmasena, Bessler, and Capps (2016) use of DAGs reveals obesity, food insecurity, and SNAP participation to be strictly endogenous. Our results find obesity to be endogenous but that food insecurity and SNAP participation are not strictly endogenous. Also similar to Dharmasena, Bessler, and Capps (2016), we find no direct causality between obesity and food insecurity. However, we find a back door path between obesity and food insecurity via income (*Obese ← household income as percent of household poverty guideline → low food security*). Conditioning on income will remove the connection between obesity and low food security in a model.

We also find multiple causal chains from food insecurity to obesity. For example, from food insecurity to fair or poor health status to obesity. This means that being food insecure causes an individual to be more likely to have a fair or poor health status which causes an individual to be more likely to be obese. The connection between fair or poor health status and obesity is quite strong as indicated in table 3.4. The appearance of this causal chain means that researchers need to be careful, as conditioning on health status will block this path between food insecurity and obesity.

Gundersen et al. (2014) model food insecurity with income, race, and unemployment at a state level. They find that the unemployed and those in poverty are more likely to be food insecure. No link was found between the Black population and food insecurity but a negative association was found with a state's Hispanic population. Nord et al. (2010) find that households headed by a Black, Hispanic, or less educated individuals are all more likely to be food insecure.

Figure 3.4 shows paths between poverty, race and food insecurity. We find that Hispanic individuals are more likely to be food insecure. There is also a direct path between the percent of poverty level and food insecurity and a path between college education and food insecurity. An income above the poverty level and being a college graduate mean that an individual is less likely to be food insecure. The college degree effect is much larger than the income effect (i.e., a college degree reduces the chance of being food insecure the same amount as an extra amount of income equivalent to 188% of the poverty level).

We find a causal chain from Black to SNAP participation to low food security. Black individuals are more likely to be SNAP participants, which then leads to a higher level of food insecurity. Both of the effects in this chain are relatively strong compared to other connections in table 3.4. Asian individuals are less likely to be SNAP participants, which would in turn reduce the level of food insecurity. These results mean that if SNAP participation is included in a model with this race variable and food insecurity, the path between race and food insecurity will be blocked by the inclusion of SNAP participation. This supports the results from Dharmasena, Bessler, and Capps (2016).

Gundersen and Ziliak (2015) report that food insecure children are more likely to report being in fair or poor health and that food insecurity is generally negatively associated with health. Our DAG shows similar results. There is a direct and positive path from low food security to fair or poor health status. There is also an indirect path from food insecurity to fair or poor health status via fair or poor diet status. Thus, including diet status in a model of food insecurity and health status might block the path between food insecurity and health status.

Conclusions

The objective of this research was to use the individual and household characteristics, characteristics of the local food environment, the individual's dietary pattern, prices, health outcomes, and policy variables to estimate a complex causality structure using the National Household Food Acquisition and Purchase Survey (FoodAPS). This was in contrast to studies that consider these variables in a fragmented approach and with which we made comparison in the results section.

To accomplish this, we estimated a graphical causality structure by way of a directed acyclic graph (DAG). The DAG is generated using two algorithms: GES and LOFS R3. First, the GES algorithms is run on the data. Then, the LOFS R3 algorithm is run on the resulting structure to orient any edges that were not oriented by the GES algorithm. The DAG is generated under assumptions made by imposing a priori knowledge on the structure. From this DAG we are able to construct structural relationships and estimate partial effects for the edges, which allowed for the comparisons to other research.

We found a number of interesting results for the relationship between individual characteristics and the local food environment. Asian individuals live in areas with higher concentrations of fast food and non-fast food restaurants. White individuals live in areas with few fast food restaurants. Hispanic individuals live in areas with a higher concentration of fast food restaurants and food stores.

We also find some interesting results between individual characteristics and health outcomes. Asian individuals are less likely to be obese. College graduates are less likely to report a fair or poor diet, less likely to report a fair or poor health status, and less likely to report low food security. Hispanic individuals are more likely to report a fair or poor diet and more likely to report low food security. Those with higher incomes are less likely to report low food security and less likely to be obese.

Our price variables have mixed results in the DAG. Individual and household characteristics, local food environment, and the policy variables do not have any causal effect on the prices. This is likely due to the procedure used to correct for quality described earlier. Only one price affects any part of the dietary pattern; the price of diary positively causes the consumption of oils. Some quantities affect price in the DAG, such as the quantity of vegetables affecting the price of vegetables. The quantity of oils also affects the price of oils. However, no price directly affects its corresponding product.

In regards to the paths between poverty, race and food insecurity, we find a number of paths. We find that Hispanic individuals are more likely to be food insecure. There is also a direct path between the percent of poverty level and food insecurity and a path between college education and food insecurity. We find a causal chain from Black to SNAP participation to low food security. A similar casual chain also exists for Hispanic individuals to low food security via SNAP participation. These results mean that if SNAP participation is included in a model with these race variables and food insecurity, the path between race and food insecurity will be blocked by the inclusion of SNAP participation. This is similar to results from Dharmasena, Bessler, and Capps (2016).

A number of directions for future research are immediately apparent. Moving beyond assuming a Normal distribution and the absence of latent variables is ripe for future research. Use of an algorithm that allows for latent variables such as the FCI (Fast Causal Inference) may give further insight into the interactions of variables (Spirtes, Meek, and Richardson, 1999). Similarly, an algorithm that allows non-Gaussian error terms such as LiNGAM may be helpful (Shimizu et al., 2006). In addition, we may need to consider the choice of variables in the model. For example, perceptions of the local food environment may be more important in influencing the consumption of fruits and vegetables than actual store density measures (Lucan and Mitra, 2012).

.

CHAPTER IV

PRE-DETERMINED DEMAND AND REGULARITY CONDITIONS OF DEMAND SYSTEMS: IMPORTANCE FOR CONSUMER FOOD DEMAND AND POLICY ANALYSIS IMPLICATIONS

Governments try to influence their citizens' diets in a number of methods. One option is to influence the price of a product to encourage more consumption of the product. For example, it is possible that a 10% subsidy for low-income Americans could increase their consumption of fruits by 2.1-5.2% and vegetables by 2.1-4.9% (Dong and Lin, 2009). A 20% subsidy on healthy dishes in a university cafeteria was followed by a 6% increase in the consumption of healthy foods and a 2% decline in the consumption of less-healthy foods (Michels et al., 2008). Experiments in laboratory settings have demonstrated that a reduction in the price of certain healthier products by 10% led to an increase in the purchase of these products by 10.3% (Epstein et al., 2010).

When trying to determine the impacts of policy, it is important to account for pre-committed demand (quantity consumed with little regard to price). This seems especially important when examining consumer food demand. Importantly, to estimate the size of policy effects it is necessary to specify the correct functional form. One example of a functional form that incorporates pre-determined demand is the Generalized Almost Ideal Demand System (GAIDS). This system extends the traditional Almost Ideal Demand System (Deaton and Muellbauer, 1980a) specification in that it allows estimation of pre-committed demand components in the budget share equations.

Another important problem for consideration is how testing and imposing theoretical regularity restrictions pertains to consumer demand theory in empirical demand analysis. This problem stems from the use of specific flexible functional forms such as the AIDS or translog (Christensen et al., 1973) for which regularity properties are often violated in practice. Given that some (such as Barnett, 2002) believe theoretical regularity must guide the selection of a functional form, several studies have examined the implications of imposing theoretical regularity conditions without sacrificing flexibility to maintain the appeal of the flexible functional forms. Local curvature can be imposed so that curvature conditions are satisfied at every data point (Ryan and Wales, 1998) while maintaining consistency with neoclassical demand theory.

As a major aspect of this research is the importance of theoretical regularity conditions and pre-committed demand to flexible functional form demand analysis, we feel that the themes running through this research are similar to those of Angus Deaton's (Deaton, 2016) body of research. Deaton (2016) believes one major theme running throughout his research is "the link between measurement, behavior and policy" (p. 1221). The measurement part of our research is the use of a household survey (Nielsen Consumer Panel). These surveys are important in that they are "used to document living standards, inequality, and poverty and, beyond that, to understand behavior (Deaton, 2016, p. 1222). The behavior aspect of our research is our use of the AIDS with the household survey. In regards to the AIDS, Deaton states that "its convenience and consistency with price theory has made it a widely used tool in work that requires inference from prices to welfare, for example in tax evaluation, regulatory, or antitrust work" (Deaton, 2016, p. 1232). Finally, the policy aspect is the use of our demand analysis results to make policy inferences about the consumption of fruits and vegetables in the United States. We believe to implement correctly the theme of measurement, behavior and policy it is a requirement that behavior be modeled with special attention to pre-committed demand and theoretical regularity conditions.

The objective of this research is to examine the affect that ignoring pre-determined demand and theoretical regularity conditions will have on consumer food demand. To accomplish this we use the AIDS because of its wide use in applied policy research. We pay additional attention to regularity by testing for compliance with these conditions. We perform the empirical analysis using Nielsen Homescan Consumer Panel data. We create a monthly time series of a representative U.S. consumer's purchases of fresh fruit, fresh vegetables and beans, frozen fruit, frozen vegetables and beans, canned fruit, canned vegetables and beans for the years 2004 through 2014. This research estimates the presence and levels of pre-committed demand. If pre-committed demand is present, then models that do not account for this are incorrectly specified. Further, the results are used to determine the effect of a 20% subsidy on canned, fresh, and frozen fruits and vegetables.

Our main findings can be briefly summarized as follows: In terms of satisfying regularity conditions, both models satisfy positivity, the AIDS with pre-commitments performs slightly better in terms of satisfying monotonicity, and both models fail to satisfy local curvature. Another important result from this study is not only the need to account for pre-commitments, but also the need to account for the consumer's timing of pre-commitments. The result from a proposed subsidy further reinforce the importance of accounting for pre-commitments. For a 20% subsidy applied to all products, the AIDS with no pre-commitments predicts a total increase of fruit and vegetable consumption of 709.9 grams per month higher than AIDS with pre-commitments with the same subsidy. For a 20% subsidy applied to only fresh fruit and vegetables, AIDS with no pre-commitments predicts a total increase of fruit and vegetable consumption of 302.5 grams per month higher than AIDS with pre-commitments with the same subsidy.

The remainder of this paper proceeds as follows. In the next section, we discuss the existing literature pertaining to pre-committed demand and produce subsidies. Then, we specify the model and outline the estimation methodology. Next, we give a detailed description of the data. Then, we discuss and present the results. Finally, we summarize the results, discuss relevance to policy application, and suggest directions for further research.

Literature Review

Since Deaton and Muellbauer's (1980a) introduction of the AIDS, the model continues to be widely applied in its many forms. Different versions of the AIDS have been used to model the effects of taxes on soft drinks (Dharmasena and Capps, 2012; Lin et al., 2011; Zhen et al., 2011) and the effects of advertising on the consumption of various beverages (Zheng and Kaiser, 2008). The AIDS has also been used in food demand to model the nutritional impacts of rising food prices in China (Zheng and Henneberry, 2012), and household food demand in Tanzania (Abdulai and Aubert, 2004). The AIDS has also been used to model non-food commodities such as the demand for gasoline (Chang and Serletis, 2013) and demand for oil, coal, and natural gas (Rowland, Mjelde, and Dharmasena, 2017). These are only a few applications of different versions of the AIDS in the extant literature.

Models incorporating pre-commitments have been estimated for a number of food demand studies, most notable the demand for meat. Tonsor and Marsh (2007) conduct an analysis of pre-committed meat and fish demand by U.S. and Japanese households using the GAIDS. They find that U.S. consumers have pre-committed demand for beef and pork while Japanese consumers have pre-committed demand for beef and fish. Their results also indicate that the GAIDS performs better than the AIDS based on in-sample and out-of-sample forecasting performance. Piggot and Marsh (2004) use the GAIDS and find pre-committed levels of meat consumption of 15.2 pounds for beef, 7.3 pounds for pork, and 10.4 pounds for poultry per quarter. These pre-commitments are impacted by seasonal factors and time trends.

Hovhannisyana and Gould (2011) examine food demand and its dynamics for 11 commodities in urban China based on household-level expenditure data for 1995 and 2003 with the GQAIDS model (quadratic form of GAIDS). The authors find no pre-committed demand in 1995 but find some level of pre-commitment for fine grains in 2003. This implies that the average Chinese household has incorporated elements of Western diet into traditional Chinese diet over time. Hovhannisyana and Gould (2014) use the GQAIDS model again with provincial-level Chinese panel data from 2002 to 2010. The results again indicated that a GQAIDS outperforms a QAIDS. These studies incorporating pre-commitments did not impose local curvature conditions in the method of Ryan and Wales (1998).

Rowland, Mjelde, and Dhamasena (2017) look at pre-commitments in aggregate energy demand in the U.S. The authors find that the level of pre-commitments range from 60% for natural gas to 87% for oil. These results are used to examine a policy that reduces oil consumption from 7.6 billion barrels to 7 billion barrels. The AIDS predicts that a 10.5% increase in price is necessary for this reduction in oil consumption versus a 44.9% increase in price for the GAIDS. The authors applied local curvature to the AIDS but were not given sensible statistically significant results.

A number of studies have been written examining methods to increase the consumption of fruits and vegetables. Dong and Lin (2009) estimate that a 10-percent subsidy would encourage low-income Americans to increase their consumption of fruits by 2.1-5.2% and vegetables by 2.1-4.9%. Klerman, Bartlett, Wilde, and Olsho (2014) study the effects of the USDA Healthy Incentives Pilot, which provided a 30% incentive for purchases of certain fruits and vegetables. These authors find that participants had a 24-percent higher intake of these fruits and vegetables compared to those in the control group. Lin, Yen, Dong, and Smallwood (2010) find that a 10% price subsidy for U.S. Supplemental Nutrition Assistance Program (SNAP) recipients focused on fruits and vegetables is predicted to increase at-home consumption of vegetables would increase from 0.94 to 1 cup (6% increase) and fruits from 0.38 to 0.42 cup (11% increase).

Waterlander et al. (2012) use a sample in the Netherlands and conduct an online experiment on shopping behavior. The authors find that a 25% discount on the total amount of fruit and vegetables purchased would lead to a 25% increase fruits and vegetables purchase. Nnoaham et al. (2009) estimate that for a United Kingdom sample that a 17.5% subsidy along with a tax on less healthy food would lead to a 5% increase in fruit and vegetable consumption. A 20% subsidy on healthy dishes in a university cafeteria was followed by a 6% increase in the consumption of healthy foods and a 2% decline in the consumption of less-healthy foods (Michels et al., 2008). None of these policy studies explicitly consider pre-determined demand.

In their work introducing a method to impose local curvature conditions on flexible demand systems, Ryan and Wales (1998) apply their method to AIDS (Deaton and Muellbauer, 1980a), normalized quadratic (Diewert and Wales, 1988), and the linear translog. Other authors have applied this method to other flexible systems including the generalized Leontief model (Serletis and Shahmoradi, 2007) and the quadratic AIDS (Chang and Serletis, 2012).

Empirical Model and Estimation Procedure

The GAIDS is an extension of the traditional AIDS specification of Deaton and Muellbauer (1980a). Bollino (1987) generalizes the AIDS by incorporating the pre-committed expenditures into the total expenditure leading to the GAIDS. The indirect utility function underlying the GAIDS is given by

$$\ln V = \frac{ln(s) - ln(P)}{b(p)}. \tag{4.1}$$

For this indirect utility function, *s* is the supernumerary expenditure (expenditure which is not affected by price) and is defined as

$$\mathrm{s} = \mathrm{m} - \textstyle\sum_{i=1}^{n} c_i p_i, \tag{4.2}$$

where m is the total expenditure of the commodities in the study, c_i are the parameters of pre-commitment to be estimated, and p_i is the price of the ith commodity. The total pre-committed expenditure is defined as $\sum_{i=1}^{n} c_i p_i$. Roy's identity is used on the indirect utility function to obtain Marshallian demand functions. These demand functions are re-arranged in budget share form to get the budget shares for GAIDS:

$$\mathrm{w_i} = \frac{c_i p_i}{m} + \frac{s}{m}\left(\alpha_i + \sum_{j=1}^{n} \gamma_{ij} \ln\left(p_j\right) + \beta \ln\left(\frac{s}{P}\right)\right). \tag{4.3}$$

In this specification $\mathrm{w_i}$ is the budget share for the ith commodity, p_i is the price of the ith commodity, α, γ, β are parameters to estimate, ln (P) is the translog price aggregator function $[\ln(P) = \alpha_0 + \sum_{j=1}^{n} \alpha_j \ln\left(p_j\right) + 0.5 \sum_{j=1}^{n} \sum_{i=1}^{n} \gamma_{ij} \ln\left(p_j\right) \ln(p_i)]$. One will notice that this budget share reduces to the AIDS form when all the $c_i = 0$ in equation (4.3). Thus if there is no pre-committed quantities found in the estimation process, then the model reduces to AIDS. For computational ease, Stone's price index is used, as suggested by Deaton and Muellbauer (1980b, p. 76), so that a Linear Approximate Generalized Almost Ideal Demand System (LA/GAIDS) will be estimated in this paper.

We include the following restrictions to satisfy demand theory of adding up: $\sum_i \alpha_i = 1$, $\sum_i \beta_i = 0$, $\sum_i \gamma_{ij} = 0$, homogeneity: $\sum_j \gamma_{ij} = 0$, and symmetry: $\gamma_{ij} = \gamma_{ji}$. It is common to include these restrictions even though they are often rejected in household or market level empirical demand estimation. Adding up is used to recover parameters for the equation that must be dropped in the estimation stage to avoid a singularity in the error covariance matrix.

One may introduce seasonality and trend variables into the budget share equations through translating of the pre-committed quantities (Pollak and Wales, 1981):

$$\tilde{c}_i = c_{i0} + \sum_{j=1}^{3} c_{ij} Q_{ij} + c_{it} * Trend_i. \tag{4.4}$$

For equation (4.4), c_{ij} are the new pre-commitment coefficients to estimate, Q_{ij} is a quarterly dummy variable, and $Trend_i$ is a linear trend variable. Introducing seasonality and trend variables into the GAIDS via the pre-committed term guarantees the invariance of elasticities to the scale of data (Alston et al., 2001). This translating will also allow examination into how the level of pre-commitments to vary over time.

There is a need to account for serial correlation as well as the possibility of contemporaneous cross equation correlation of the error terms. To account for this we use the method proposed by Berndt and Savin (1975). Given that w_{it} is the budget share, x_{it} is the list of independent variables, and v_{it} is the serially correlated disturbance term our budget share equation can be written in the following form,

$$w_{it} = f(x_{it}, \beta) + v_{it}. \tag{4.5}$$

Then given that ρ is the autocorrelation coefficient and ε_{it} is white noise disturbance term, we can write the general form of the estimating equation as,

$$w_{it} = \sum_k \rho_k w_{it-k} + f(x_{it}, \beta) - \sum_k \rho_k f(x_{it-k}, \beta) + \varepsilon_{it}, \tag{4.6}$$

Where *k* is the order of autocorrelation.

We test the three regularity conditions of positivity, monotonicity, and curvature as suggested by Barnett and Serletis (2008). These conditions are tested at each of the observations. Positivity is satisfied if the indirect utility function is positive at an observation. That is we check if $\widehat{\ln V} > 0$ for all observations. Monotonicity is satisfied if each component of the first gradient of the indirect utility function is negative when using prices normalized over expenditure. This is checked by ensuring that $\nabla \widehat{\ln V} < 0$

for all observations using normalized prices. Curvature requires monotonicity to hold and the Allen elasticity of substitution matrix to be negative semidefinite. Allen elasticities (ae_{ij}) are calculated as,

$$ae_{ij} = \frac{s_{ij}}{w_j}, \tag{4.7}$$

where s_{ij} is the compensated elasticity (defined later) divided by the expenditure share for the *jth* good.

The procedure to impose the local curvature condition on flexible functional forms is outlined in Ryan and Wales (1998) and Barnett and Serletis (2008). At the point of approximation, the *n x n* Slutsky matrix can be written as $\boldsymbol{S} = \boldsymbol{B} + \boldsymbol{C}$, where $\boldsymbol{B}$ is an *n x n* symmetric matrix with the same number of elements as the Slutsky matrix and $\boldsymbol{C}$ is an *n x n* matrix containing elements that are functions of other elements in the system. Curvature is imposed by replacing $\boldsymbol{S}$ with $-\boldsymbol{KK}'$ ($\boldsymbol{K}$ is lower triangular matrix) and then solving for $\boldsymbol{B}$ to get $\boldsymbol{B} = -\boldsymbol{KK}' - \boldsymbol{C}'$**.** The model is then reparametrized by estimating the parameters in $\boldsymbol{K}$ and $\boldsymbol{C}$ (not $\boldsymbol{B}$ and $\boldsymbol{C}$**)**. This procedure ensures that the matrix is negative semidefinite at any data point.

We estimate uncompensated and compensated elasticity estimates for the AIDS using one of the recommended forms from Alston, Foster, and Green (1994). The elasticities are calculated at the means of the data. The formula for the AIDS uncompensated elasticity is

$$e_{ij}^{A} = -\delta_{ij} + \frac{\gamma_{ij}}{w_i} - \frac{\beta_i}{w_i} w_j, \tag{4.8}$$

the formula for the AIDS compensated elasticity is

$$s_{ij}^{A} = -\delta_{ij} + \frac{\gamma_{ij}}{w_i} + w_j, \tag{4.9}$$

and the formula for the AIDS expenditure elasticity is

$$\varepsilon_{ij}^{A} = 1 + \frac{\beta_i}{w_i}. \tag{4.10}$$

The elasticity equations for the GAIDS are taken from Tonsor and Marsh (2007). The formula for the GAIDS uncompensated elasticity is

$$e_{ij}^{G} = -\delta_{ij} + \left(\frac{1}{mw_i}\right)\left[c_i p_i \left(1 - \frac{p_i(q_i - c_i)}{\mathrm{s}}\right) + s\left(\gamma_{ij} - \beta_i \left(\frac{c_i p_i}{\mathrm{s}} + \alpha_\mathrm{i} + \sum_{j=1}^{n} \gamma_{ij} \ln(p_j)\right)\right)\right], \tag{4.11}$$

the formula for the GAIDS expenditure elasticity is

$$\varepsilon_{ij}^{G} = 1 + \left(\frac{1}{w_i}\right)\left(\beta_i + \frac{1}{m}\right)(-c_i p_i + sw_i\,), \tag{4.12}$$

and the formula for the GAIDS compensated elasticity is

$$s_{ij}^{G} = e_{ij} + w_j \varepsilon_{ij}. \tag{4.13}$$

For the above elasticity equations δ_{ij} is the Kronecker delta,

$$\delta_{ij} = \begin{cases} 1 \; if \; i = j \\ 0 \; if \; i \neq j \end{cases}, \tag{4.14}$$

γ_{ij}, β_i, and α_i are the parameters to be estimated, w_i is the expenditure share for the commodity *i*, and w_j is the expenditure share for the commodity *j*, *m* is the total expenditure, c_i is the pre-commitment for commodity *i*, p_i is the price for commodity *i*, and *s* is the supernumerary expenditure defined in equation (4.2). For GAIDS, the elasticities are the changes in the discretionary level of consumption. The discretionary level of consumption is the total consumption minus the pre-committed amount of consumption for a given commodity. The discretionary level is the portion of total consumption that is affected by changes in price. The non-discretionary part of consumption is also referred to as the level of pre-commitment and is not affected by changes in price. For AIDS since there is no pre-commitment, then the total consumption for a commodity is the same as the discretionary consumption for that commodity. Thus, we expect to find that the calculated elasticities for GAIDS will be more elastic.

In order to deal with possible endogeneity in the total expenditure variable we utilize a procedure from Capps et al. (1994) and implemented in Dharmasena and Capps (2012). For this, predicted values of total expenditure are used as an instrument for observed total expenditure. Predicted values of total expenditure are obtained by regressing observed total expenditures on the prices of each product and income. This regression also includes a first order and twelfth order autoregressive error terms to account for serial correlation. This predicted value is used in the estimation procedure in place of the observed value.

Unit values (proxy for prices) are calculated by taking the total expenditure in a category and dividing this by the total weight (grams) purchased. Einav, Leibtag and Nevo (2010) find that Homescan prices are measured with errors. The use of unit values is also likely to create bias in the form of measurement error. It is possible the commodity aggregates are endogenous to the choice of quality. As the representative amount in our data is based on household demographics, which are changing over time, it is necessary to account for this endogeneity. We utilize the procedure described by Cox and Wohlgenant (1986) to correct for endogeneity in prices.

For this procedure, we regress the difference between the unit price and the mean unit price for each category on a number of household demographics.

$$p_i^u - \bar{p}_i^u = \sum_j \beta_{ij} D_{ij} + v_i \tag{4.15}$$

For equation (4.15), p_i^u is the unit price for a commodity in a given month, $\bar{p}_i^u$ is the mean unit price across all months, β_{ij} is a set of coefficients to be estimated, D_{ij} is a vector of household characteristics, and v_i is the error term. The demographics used in the regression are the characteristics of a representative household in the U.S. for each month in the sample. Since the representative individual's characteristics change over time (following trends in the U.S. as a whole), this adjustment is necessary to account for a possibility that the quality of the commodities purchased may change over time. We include income, household size, and dummy variable indicating the region of the country. In order to get the quality-adjusted price, we used the estimated coefficients from equation (4.15) and then calculate the following,

$$\hat{p}_i = p_i^u - \sum_j \hat{\beta}_{ij} D_{ij}, \tag{4.16}$$

where $\hat{p}_i$ are the prices to be used in the estimation in place of the observed unit prices, p_i^u is the unit price for a commodity, $\hat{\beta}_{ij}$ is the set of estimated coefficients, and D_{ij} is a vector of household characteristics.

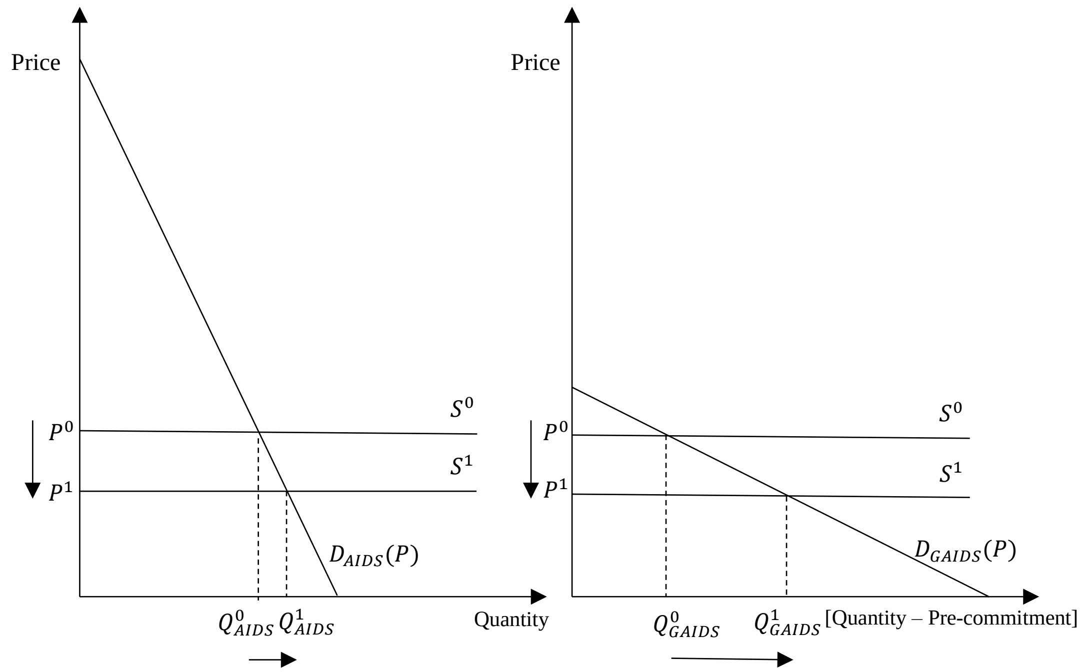
Serial correlation is corrected using the procedure suggested by Berndt and Savin (1975). A second order correction is used. Along with the corrected expenditure and price from above, the estimated form adjusted for serial correlation of the LA/GAIDS equation is

$$\mathrm{w_{it}} = \rho_1 \mathrm{w_{it-1}} - \rho_2 \mathrm{w_{it-2}} + \left[\frac{\tilde{c}_i p_i}{m} + \frac{s}{m}\left(\alpha_i + \sum_{j=1}^{n} \gamma_{ij} \ln(p_j) + \beta \ln\left(\frac{s}{P_t}\right)\right)\right] - \rho_1 \left[\frac{\tilde{c}_i p_i}{m} + \frac{s}{m}\left(\alpha_i + \sum_{j=1}^{n} \gamma_{ij} \ln(p_{jt-1}) + \beta \ln\left(\frac{s}{P_{t-1}}\right)\right)\right] - \rho_2 \left[\frac{\tilde{c}_i p_i}{m} + \frac{s}{m}\left(\alpha_i + \sum_{j=1}^{n} \gamma_{ij} \ln(p_{jt-2}) + \beta \ln\left(\frac{s}{P_{t-2}}\right)\right)\right] + e_{it}, \tag{4.17}$$

where $\mathrm{s} = \mathrm{m} - \sum_{i=1}^{n} c_i p_i$, $\widetilde{c_i} = c_{i0} + \sum_{j=1}^{3} c_{ij} Q_{ij} + c_{it} * Trend_i, P_t = \sum w_j lnp_j$, and e_{it} is the error term. The LA/AIDS adjusted for serial correlation is equation (4.17) where all c_i equal zero. The equations are estimated using the iterated seemingly unrelated regression procedure in the SAS proc model command (SAS Institute Inc., 2014). Elasticities are calculated using equations (4.8) through (4.13).

Then, we conduct a partial equilibrium analysis in order to find the effects of a proposed subsidy for both models. This procedure assumes that any increase in demand will be met by an increase in supply at the current price. This is a situation with a relatively inelastic demand curve and a perfectly elastic supply curve, which means a 100% pass through of subsidy effect to consumers. In effect, the consumers will be facing a lower price when shopping for fruits and vegetables.

Figure 4.1. Partial Equilibrium Effects of a Subsidy on Fruits and Vegetables



Note: This figure illustrates the change in quantity from a decrease in price (a subsidy). The more elastic demand curve represents the GAIDS while the more inelastic demand curve represents the AIDS. Notice that the horizontal axis on the AIDS graph is quantity while the horizontal axis on the GAIDS is quantity minus the level pre-commitments. A given decrease in price will lead to a larger quantity change for the GAIDS, as the demand curve is more elastic. The total change may be lower for the GAIDS if there is a high level of pre-commitment for the commodity.
Source: Produced by author.

Figure 4.1 shows a general representation of these assumptions. Figure 4.1 includes a demand curve in the right panel that more elastic than the demand curve in the left panel. This is similar to the situation we are facing in that the elasticity estimates from the GAIDS are expected to be more elastic than the AIDS. The more elastic demand curve represents the GAIDS while the relatively more inelastic demand curve represents the AIDS. This figure illustrates the change in quantity from a decrease in price (a subsidy). It is important to keep in mind that the subsidy will only affect the discretionary portion of demand. For the AIDS the discretionary portion is the entire quantity while for the GAIDS the discretionary portion is the quantity minus the level of pre-commitment. A given decrease in price is predicted to lead to a larger percent change in quantity for the GAIDS because the demand curve is more elastic. The total change may be lower for the GAIDS if there is a high level of pre-commitment for the commodity. Appendix E explores the condition under which the GAIDS will have a larger quantity change than the AIDS.

Two different scenarios are analyzed for the 20% subsidy: on all fruits and vegetables and only on fresh fruits and vegetables. We begin by finding a baseline discretionary level of purchase for the AIDS as the average amount purchased for the last 12 months of data for each respective commodity. Then for each commodity, we increase or decrease this discretionary quantity by the corresponding own and cross price uncompensated elasticities by assuming a 20% decrease in the price faced by consumers (the subsidy).

Similarly, for the GAIDS we begin by finding the baseline discretionary level. The baseline level will equal the average amount purchased for the last 12 months minus the average level of the estimated pre-commitments for the last 12 months. The discretionary portion for the GAIDS will be less than for the AIDS due to the inclusion of pre-commitments, which are non-discretionary. For each commodity, we increase or decrease this discretionary quantity by the corresponding own and cross price uncompensated elasticities by assuming a 20% decrease in the price faced by consumers (the subsidy). For the GAIDS the new discretionary amount is added back to the pre-committed amount to get the new estimated total level of purchase for the consumer.

Data

Data are obtained from Nielsen Homescan Consumer Panel.[11] We create a monthly time series of a representative U.S. consumer's purchases for the years 2004 through 2014 (132 months) by utilizing the sampling weights, as explained below. Each participating household is given a scanner to read UPCs from products purchased at stores. Nielsen matches the scanned UPC with products characteristics in their database. The household is also asked to enter quantity, expenditure, and any coupon information about the products. The food products selected for study are fresh fruit, fresh vegetables and beans, frozen fruit, frozen vegetables and beans, canned fruit, canned vegetables and beans. The sampling weights are key in order to calculate a representative level of consumption as some households are more likely to be selected based on demographic characteristics than compared to the general population. The sampling weights must be used to correct for this sample bias. The weights were derived by re-balancing the raw panel so that the weighted panel will match ten standard demographic variables (household size, income, race, etc.) for the U.S. (Kilts Center for Marketing, 2013). The sum of the weights is the total number of households in the U.S.

Quantities and expenditures are totaled across a household sampled in the panel for a given month. These totals are multiplied by the sampling weight for that household.

[11] Data from The Nielsen Company (US), LLC and marketing databases provided by the Kilts Center for Marketing Center at The University of Chicago Booth School of Business.

Then we total across all households for a given month. This total then divided by the sum of sampling weights for that month to create an approximation of monthly per capita purchase in grams consumed for the category and per capita expenditures. Prices (unit values) are then calculated by dividing total expenditure by total quantity purchased. Price are quality adjusted using the procedure mentioned earlier using equation (4.16).

Table 4.1. Summary Statistics of Monthly Data, January 2004 through December 2014

Variable	Mean	Std. Dev	Minimum	Maximum
Quantity purchased per month (grams/capita)				
Canned fruit	226.51	72.38	134.83	492.96
Fresh fruit	1368.38	178.60	976.76	1864.40
Frozen fruit	32.32	6.26	22.03	51.48
Canned vegetables	647.86	130.71	442.58	976.14
Fresh vegetables	1336.14	137.66	1001.00	1706.14
Frozen Vegetables	390.01	42.21	305.43	497.36
Unit price (cents/gram)				
Canned fruit	0.11	0.01	0.10	0.13
Fresh fruit	0.11	0.01	0.09	0.14
Frozen fruit	0.26	0.01	0.23	0.29
Canned vegetables	0.09	0.01	0.08	0.10
Fresh vegetables	0.12	0.01	0.09	0.15
Frozen Vegetables	0.15	0.01	0.14	0.16
Expenditure share				
Canned fruit	0.05	0.02	0.03	0.11
Fresh fruit	0.33	0.03	0.27	0.43
Frozen fruit	0.02	0.00	0.01	0.02
Canned vegetables	0.12	0.02	0.08	0.17
Fresh vegetables	0.34	0.02	0.30	0.38
Frozen Vegetables	0.13	0.01	0.09	0.15

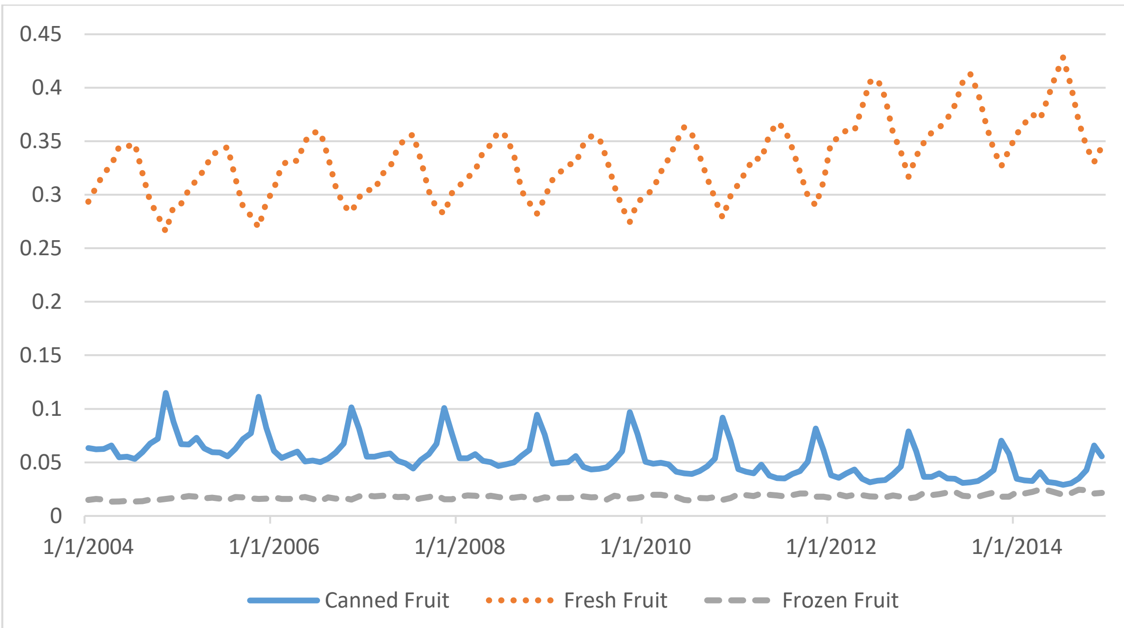
Source: Calculated by author.

Table 4.1 shows the summary statistics for the variables used in the estimation. The quantity purchased is expressed as grams per capita for a given month. Fresh fruit and fresh vegetables comprise the largest quantity purchased with 1368.38 (g/capita) and 1336.14 (g/capita) respectively. Canned fruit and frozen fruit are the two smallest categories purchased with 226.51 (g/month) and 32.32 (g/capita) respectively. In the middle are frozen vegetables at 390.01 (g/capita) and canned vegetables at 647.86 (g/capita).

The prices presented in table 4.1 are the unit prices prior to the quality adjustment. On average, the unit price is the highest for frozen fruit at 0.26 cents per gram and the lowest for canned vegetables at 0.09 cents per gram. Canned fruit and fresh fruit has a similar mean at 0.11 cents per gram. Fresh vegetables and frozen vegetables comprise the remainder at 0.12 cents per gram and 0.15 cents per gram respectively. Fresh fruit and fresh vegetables have the highest expenditure shares at 0.33 and 0.34 respectively. Frozen fruit has the smallest expenditure share at 0.02 followed by canned fruit at 0.05. The remaining expenditure shares include canned vegetables at 0.12 and frozen vegetables at 0.13.

Figure 4.2 is a graph of the monthly expenditure shares of the representative household for fruit products. Fresh and canned fruit show a high degree of seasonality. The expenditure share for fresh fruit dips in October, November, and December with the share of canned fruit increasing to compensate for the decline. In contrast, the expenditure share for fresh fruit peaks in May, June, and July, which coincides with the minimum expenditure share for canned fruit. Frozen fruit shows little noticeable trend or seasonality. The expenditure share for fresh fruit shows a slight upward trend near the end with canned fruit showing a slight decrease. These behaviors show a need to compensate for trend and seasonality in the modeling for these fruit products.

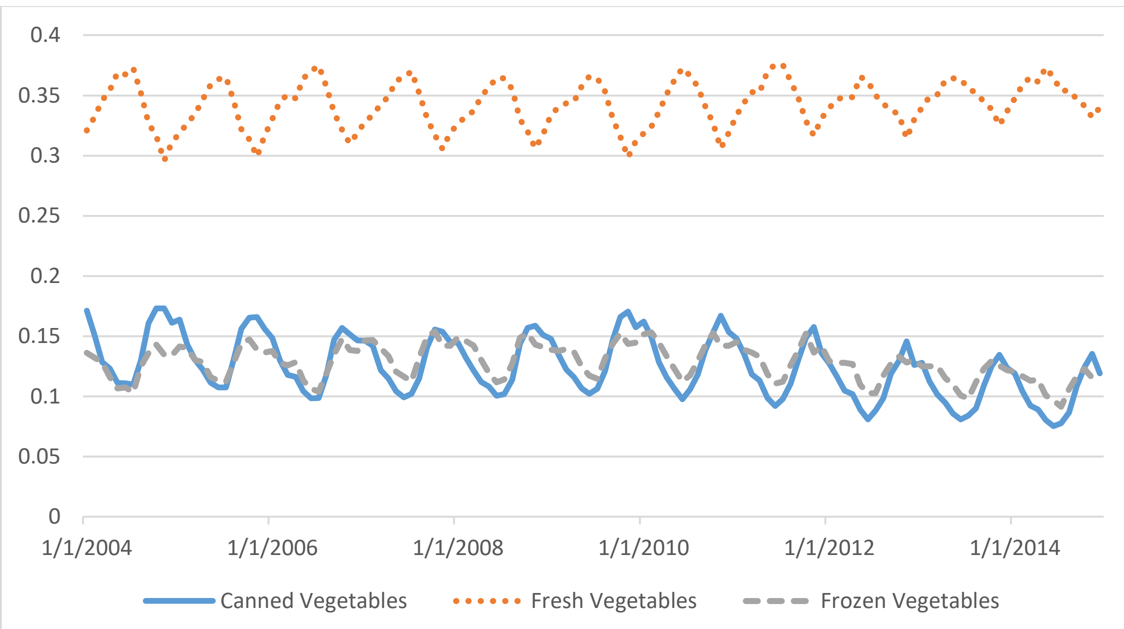
Figure 4.2. Expenditure Shares for Fruit Products, January 2004 through December 2014



Source: Produced by author.

Figure 4.3 is a graph of the monthly expenditure shares of the representative household for vegetable products. All three types show a high degree of seasonality. In a similar manner to the fruit products, consumption for fresh vegetables peaks in May, June, and July and reaches it minimum expenditure share in October, November, and December. The inverse is true for both canned and frozen vegetables as the expenditure share peaks in October, November, and December while reaching a minimum expenditure share in May, June, and July. The pattern where a decrease in the expenditure share for fresh vegetables products in compensated by an increase in the canned vegetables is similar to what is seen with the fruit products. There does not appear to be a noticeable trend in the fresh vegetable products, upon visual examination. The frozen and canned vegetables show a small decline near the end of the sample. This again demonstrates a need to account for seasonality and trend in the variables during modeling.

Figure 4.3. Expenditure Shares for Vegetable Products, January 2004 through December 2014



Source: Produced by author.

In order to check the validity of our constructed sample, we compared our sample to overall trends in the United States given by the UDSA's Food Availability (Per Capita) Data System (USDA, ERS, 2017a). Food loss represents the edible amount of food, postharvest, that is available for human consumption but is not consumed for any reason. Food loss includes cooking loss, loss from mold or pests, and food waste. This data is helpful for the comparison because the USDA data serves as an indirect measure of trends in food purchases. Both data series provide an indication of whether Americans, on average, are consuming more or less of various foods over time. The correlations when we compare our data to the USDA's per capita availability adjust for loss are 0.90 for fresh fruit, 0.84 for canned vegetables, 0.82 for canned fruit, 0.46 for frozen fruit, -0.31 for fresh vegetables, and -0.57 for frozen vegetables. The two negative are explained as our sample increasing their purchase of fresh vegetables and frozen vegetables relative to a decreasing availability for the USDA data. In line with the trends in our data, another national study of produce consumption indicates that the consumption of fresh and frozen vegetables has increased over the period from 2004 to 2014 (Produce for Better Health Foundation. 2015, p. 10). These results indicate that our sample constructed from the Nielsen Homescan Consumer Panel follows national trends in produce consumption.

Table 4.2. Parameter Estimates for the LA/AIDS and LA/GAIDS Models

Parameter	LA/AIDS Estimate	LA/AIDS p-value	LA/GAIDS Estimate	LA/GAIDS p-value
c_{Fcn0}			-280.6560	0.1299
c_{Ffr0}			**-777.9870**	**0.0245**
c_{Ffz0}			-1.4264	0.8933
c_{Vcn0}			242.9835	0.2538
c_{Vfr0}			**-998.0020**	**0.0003**
c_{Vfz0}			81.6642	0.3019
α_{Fcn}	**-0.5359**	**0.0210**	0.0268	0.9210
α_{Ffr}	0.0321	0.9102	-0.2592	0.4552
α_{Ffz}	0.0295	0.4303	-0.0093	0.8237
α_{Vcn}	0.2722	0.1889	-0.1438	0.5882
α_{Vfr}	**0.4134**	**0.0271**	**0.9569**	**0.0001**
α_{Vfz}	**0.7888**	**<.0001**	**0.4286**	**0.0253**
β_{Fcn}	**0.0798**	**0.0053**	0.0145	0.6347
β_{Ffr}	0.0273	0.4337	**0.0653**	**0.0985**
β_{Ffz}	-0.0029	0.5326	0.0019	0.6989
β_{Vcn}	-0.0159	0.5309	0.0223	0.4432
β_{Vfr}	-0.0115	0.6123	**-0.0648**	**0.0190**
β_{Vfz}	**-0.0768**	**0.0003**	**-0.0392**	**0.0733**
γ_{FcnFcn}	0.0294	0.1954	**0.0656**	**0.0102**
γ_{FcnFfr}	**-0.0805**	**<.0001**	**-0.1143**	**<.0001**
γ_{FcnFfz}	-0.0059	0.2246	**-0.0093**	**0.0931**
γ_{FcnVcn}	**0.0474**	**0.0013**	**0.0759**	**<.0001**
γ_{FcnVfr}	0.0185	0.1429	-0.0132	0.3196
γ_{FcnVfz}	-0.0090	0.6117	-0.0047	0.7864
γ_{FfrFfr}	**0.0892**	**0.0004**	0.0405	0.2578
γ_{FfrFfz}	0.0017	0.7555	0.0061	0.2702
γ_{FfrVcn}	**-0.0328**	**0.0480**	-0.0009	0.9565
γ_{FfrVfr}	**0.0315**	**0.0537**	0.0410	0.1750
γ_{FfrVfz}	-0.0092	0.5845	**0.0276**	**0.0907**
γ_{FfzFfz}	**0.0120**	**0.0100**	0.0060	0.3154
γ_{FfzVcn}	-0.0004	0.9353	0.0049	0.3184
γ_{FfzVfr}	-0.0025	0.5914	-0.0030	0.5609
γ_{FfzVfz}	-0.0050	0.3900	-0.0047	0.3985
γ_{VcnVcn}	**-0.0680**	**<.0001**	**-0.1167**	**0.0015**
γ_{VcnVfr}	**-0.0202**	**0.0861**	-0.0130	0.3318
γ_{VcnVfz}	**0.0739**	**<.0001**	**0.0500**	**0.0024**
γ_{VfrVfr}	**0.0388**	**0.0074**	**0.0501**	**0.0484**
γ_{VfrVfz}	**-0.0662**	**<.0001**	**-0.0619**	**<.0001**

(continued)

Table 4.2. ***(continued)***

	LA/AIDS		LA/GAIDS	
Parameter	Estimate	p-value	Estimate	p-value
γ_{VfzFcn}	-0.0090	0.6117	-0.0047	0.7864
γ_{VfzFfr}	-0.0092	0.5845	**0.0276**	**0.0908**
γ_{VfzFfz}	-0.0050	0.3900	-0.0047	0.3986
γ_{VfzVcn}	**0.0739**	**<.0001**	**0.0500**	**0.0024**
γ_{VfzVfr}	**-0.0662**	**<.0001**	**-0.0619**	**<.0001**
γ_{VfzVfz}	0.0155	0.4591	-0.0062	0.7951
Q_{Fcn1}	**-0.0158**	**<.0001**	**-202.6590**	**0.0393**
Q_{Ffr1}	**0.0121**	**0.0002**	53.6313	0.8330
Q_{Ffz1}	**0.0010**	**0.0142**	4.3225	0.2656
Q_{Vcn1}	**-0.0131**	**<.0001**	**-221.9970**	**0.0182**
Q_{Vfr1}	**0.0139**	**<.0001**	85.9788	0.6749
Q_{Fcn2}	**-0.0130**	**<.0001**	**253.9815**	**0.0480**
Q_{Ffr2}	**0.0240**	**<.0001**	**1342.1420**	**<.0001**
Q_{Ffz2}	0.0004	0.2912	**17.3627**	**0.0077**
Q_{Vcn2}	**-0.0211**	**<.0001**	87.5134	0.4984
Q_{Vfr2}	**0.0189**	**<.0001**	**1235.8540**	**<.0001**
Q_{Fcn3}	**-0.0114**	**<.0001**	86.3827	0.3670
Q_{Ffr3}	**0.0107**	**0.0007**	**708.8049**	**0.0010**
Q_{Ffz3}	0.0005	0.2263	**10.3405**	**0.0197**
Q_{Vcn3}	**-0.0083**	**0.0004**	129.3558	0.1895
Q_{Vfr3}	**0.0094**	**<.0001**	**654.7915**	**0.0008**
$Trend_{Fcn}$	**-0.0002**	**<.0001**	0.3605	0.5609
$Trend_{Ffr}$	**0.0002**	**<.0001**	**7.6103**	**<.0001**
$Trend_{Ffz}$	**0.0000**	**0.0005**	**0.1785**	**<.0001**
$Trend_{Vcn}$	**-0.0002**	**<.0001**	0.0092	0.9888
$Trend_{Vfr}$	**0.0001**	**0.0006**	**5.9534**	**<.0001**
ρ_1	**0.6897**	**<.0001**	**0.6642**	**<.0001**
ρ_2	**-0.0801**	**0.0577**	-0.0220	0.5902
Objective Fn. Value		4.639		4.56
Objective Fn. Value*N		603.0		593.0
R		0.0543		0.0007
Positivity Violations		0		0
Monotonicity Violations		6		0
Curvature Violations		132		132

Note: Coefficients in bold are significant at the 10% level. The α_i, β_i, and γ_{ij} are standard AIDS terms. Q_{ij} and $Trend_i$ are quarterly seasonality and linear trend coefficients. C_{i0} is the intercept term for the translated pre-committed quantity. The serial correlation coefficients are ρ_1 and ρ_2.The abbreviations are as follows: fcn = canned fruit, ffr = fresh fruit, ffz = frozen fruit, vcn = canned vegetables, vfr = fresh vegetables, vfz = frozen vegetables. R measures the degree to which the residuals are orthogonal to the Jacobian columns, and it approaches zero as the gradient of the objective function becomes small. The objective function value is the final value of the objective function being minimized. Objective*N can be thought of as a higher order measure of sum of squared errors (SSE) and reduces to this measure for a single equation. Smaller values of these measures indicate a better fitting function.
Source: Calculated by author.

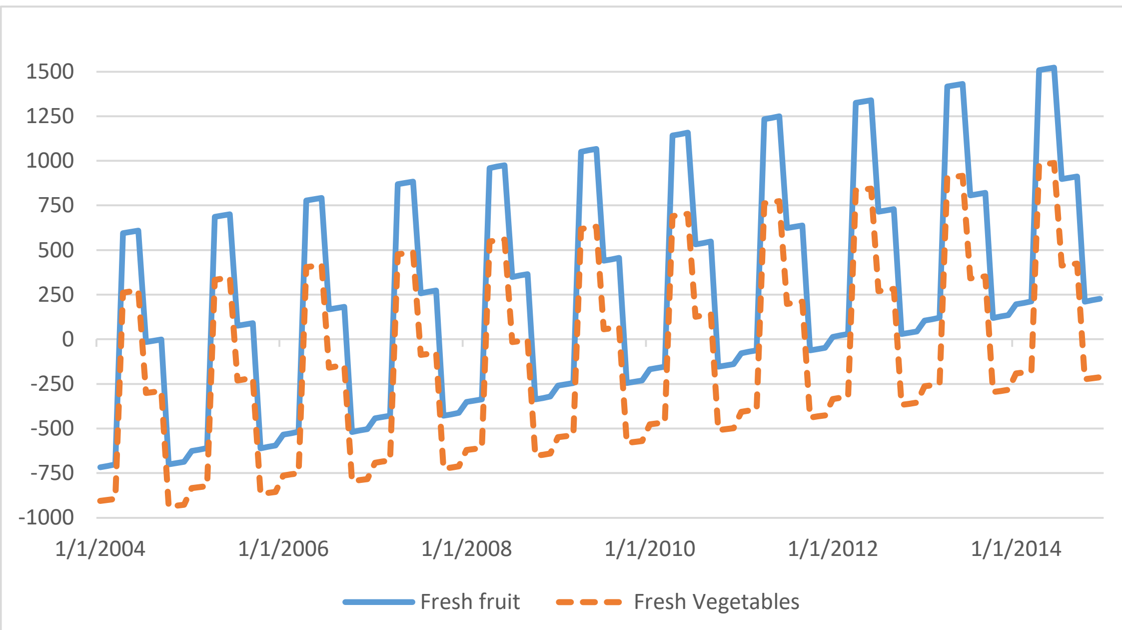
Results

Table 4.2 presents the coefficient estimates from LA/GAIDS (equation (4.17)) and the LA/AIDS, side by side. These parameters do not have an easy direct interpretation but some information can be gained from examining them. A little over half the parameters for the AIDS are significant at the 10% level while slightly under half the parameters for the GAIDS are significant at the same level. The ρ_1 parameter is large and significant for both models indicating the need to correct for first order serial correlation. The ρ_2 parameter is significant at the 10% level for the AIDS but is not significant at this level for the GAIDS.

The c_{i0} coefficients for the GAIDS model, presented in table 4.2, represent the intercept terms in equation (4.4), the translated pre-committed quantities. Further information about the level of pre-commitment can be recovered because of the translating procedure. The pre-committed amounts vary by quarter and by trend due to translating. Figure 4.4 is an example of the pre-committed amounts graph over the time horizon for fresh fruit and fresh vegetables. These two categories have the largest amounts of pre-committed demand than any other categories. In addition, figure 4.4 shows definite trend and seasonality in the pre-committed amounts of fresh fruit and fresh vegetables.

The α_i, β_i, and γ_{ij} terms from each model are directly comparable while the quarterly dummy and linear trend coefficients are not. This is due to how the seasonality was introduced into each model. The seasonality and trend for the AIDS model were introduced in a manner so that they represent the effect of trend and seasonality on the expenditure share. For the GAIDS model, the trend and seasonality were introduced through translating (equation (4.4)) so that they coefficient represent the effects on seasonality on the pre-committed quantity. That many of these quarterly seasonality and linear trend coefficients are significant confirms some of the preliminary conclusions drawn from visually examining figure 4.2 and figure 4.3.

Figure 4.4. Graph of Estimated Pre-Commitments for Fresh Fruit and Fresh Vegetables, January 2004 through December 2014



Note: Recall the pre-commitments are translated with the following equation: $\tilde{c_i} = c_{i0} + \sum_{j=1}^{3} c_{ij} D_j + c_{it} * Trend_i$. Thus, pre-commitment may vary over the sample period and within each quarter. Tonsor and Marsh (2007) explain negative estimates of pre-committed demand as a preference to shift out of pre-committed expenditure and into supernumerary expenditures (those that are not affected by price). Thus, months with negative values for pre-commitments indicated the consumer is more affected by economic factors, such as price.
Source: Produced by author.

More information about the level of pre-commitments are given in table 4.3. This table illustrates that in some months, the estimated pre-commitments are positive while in other months the pre-commitments are negative. The appearance of negative pre-commitments is counterintuitive at first but not unique as other authors have encountered this in empirical work (Tonsor and Marsh, 2007; Zheng and Henneberry, 2009; Hovhannisyana and Gould, 2011). One reason negative values appear is because regularity conditions place no restrictions on the pre-commitment parameters. Tonsor and Marsh (2007) explain a negative estimate of pre-committed demand as a marginal response of pre-committed expenditure to the price of the commodity. This means that a negative value indicates for that month that the consumer is predisposed to supernumerary expenditures instead of pre-committed expenditures. Thus in the months with negative values, the consumer is more affected by economic factors such as price and is less focused on pre-commitment.

This finding is important for policy analysis, as the consumer is more sensitive to price changes in some months relative to other months. For example, in months were the consumer has a high level of pre-commitment for fresh vegetables, any policy focusing on the price of fresh vegetables will have less affect than months where the consumer has low pre-commitments. Furthermore, it is to our understanding that the presence and implications of the estimated negative pre-commitments is an area in need of further research.

Model fit parameters are given at the end of table 4.2 for each system. The GAIDS has lower values of objective function value and R.[12] This indicates that the GAIDS fits the data better than the AIDS. A number of Wald tests are conducted on the pre-committed parameters. A joint test that the size pre-commitment intercept terms in equation (4.4) are jointly equal to zero is reject at the 0.0001 level. A joint test that the pre-committed values are equal to zero tested at each of the 132 points in the time series is rejected with the largest p-value being 0.0007. These tests further suggest that the GAIDS model is more appropriate than AIDS for the data.

Table 4.3. Summary Statistics of Pre-commitment Levels

Product	Mean Quantity Purchased (grams/capita)	Level of Pre-commitment (grams/capita) Mean	Std. dev.	Minimum	Maximum
Canned fruit	226.5	-222.26	166.11	-482.95	18.75
Fresh fruit	1368.4	254.24	618.29	-716.75	1523.05
Frozen fruit	32.3	18.45	9.32	0.36	38.43
Canned vegetables	647.9	242.31	136.24	21.00	373.52
Fresh vegetables	1336.1	-107.94	544.65	-938.47	987.98
Frozen vegetables	390.0	156.13	35.90	88.44	228.42

Note: Quantities are in grams purchased per month per capita. Recall the pre-commitments are translated with the following equation: $\widetilde{c_i} = c_{i0} + \sum_{j=1}^{3} c_{ij} D_j + c_{it} * Trend_i$. Thus, pre-commitment may vary over the sample period and by quarter. Tonsor and Marsh (2007) explain negative estimates of pre-committed demand as a preference to shift out of pre-committed expenditure and into supernumerary expenditures (those that are not affected by price). Thus, months with negative values for pre-commitments indicated the consumer is more affected by economic factors, such as price.
Source: Calculated by author.

[12] R measures the degree to which the residuals are orthogonal to the Jacobian columns, and it approaches zero as the gradient of the objective function becomes small. The objective function value is the final value of the objective function being minimized. Objective*N can be thought of as a higher order measure of sum of squared errors (SSE) and reduces to this measure for a single equation. Smaller values of these measures indicate a better fitting function.

Details of the theoretical regularity condition tests are also given at the end of table 4.2. Positivity is satisfied at every point for the LA/AIDS and the LA/GAIDS. Monotonicity is violated at six points for the LA/AIDS and at zero points for the LA/GAIDS. The inclusion of pre-commitments did lead to a better result in this case. Unfortunately, the curvature condition is violated at every point for both models. This demonstrates a need to attempt to correct for this issue.

An attempt was made to apply the curvature conditions using the method proposed by Ryan and Wales (1998). The statistical program used for this was not able to find a closed form solution. This is likely due to the complexity of the structure imposed by this method causing issues with the numerical estimation. The $\boldsymbol{KK'}$ is a *5x5* matrix (recall that one equation is dropped due to the adding up restriction or to avoid the singularity of the variance-covariance matrix) with 25 elements of potential restrictions. Using the assumption of symmetry, this implies that we are adding 15 new restrictions to our estimation. These restrictions are used to reparametrize the model, and they are more complex than the parameters they replace. For example, we need to replace γ_{VfrVfr} with $\mathrm{k}^2_{\mathrm{VfrFcn}} + \mathrm{k}^2_{\mathrm{VfrFfr}} + \mathrm{k}^2_{\mathrm{VfrFfz}} + \mathrm{k}^2_{\mathrm{VfrVcn}} + \mathrm{k}^2_{\mathrm{VfrVfr}}$. This issue may also be due to the highly aggregated nature of the data. Exploration of alternate methods to impose the curvature restriction is an important area for future research.

Table 4.4. Uncompensated Elasticity Estimates for the LA/AIDS Model

	Canned Fruit	Fresh Fruit	Frozen Fruit	Canned Vegetables	Fresh Vegetables	Frozen Vegetables	Expenditure Elasticity
Canned	-0.5369	**-1.9757**	-0.1345	**0.69154**	-0.1626	-0.3558	**2.4740**
Fruit	0.2021	**<.0001**	0.138	**0.0181**	0.5345	0.3081	**<.0001**
Fresh	**-0.2464**	**-0.759**	0.00366	**-0.1089**	0.06658	-0.0382	**1.0821**
Fruit	**<.0001**	**<.0001**	0.8275	**0.0488**	0.2284	0.4929	**<.0001**
Frozen	-0.318	0.14976	-0.3269	-0.0008	-0.0836	-0.2568	**0.8362**
Fruit	0.2332	0.6102	0.2066	0.9976	0.7584	0.4406	**0.0018**
Canned	**0.38699**	**-0.2206**	-0.0008	**-1.5295**	-0.118	**0.60908**	**0.8728**
Vegetables	**0.0012**	**0.0783**	0.9839	**<.0001**	0.2561	**<.0001**	**<.0001**
Fresh	0.05593	**0.10311**	-0.0067	-0.0547	**-0.8751**	**-0.189**	**0.9664**
Vegetables	0.1353	**0.0284**	0.6265	0.1386	**<.0001**	**<.0001**	**<.0001**
Frozen	-0.0381	0.12766	-0.0281	**0.65157**	**-0.3113**	**-0.8023**	**0.4005**
Vegetables	0.7823	0.2971	0.5393	**<.0001**	**0.0026**	**<.0001**	**0.0148**

Note: The table presents estimated uncompensated elasticities from the LA/AIDS. Elasticities in bold are significant at the 10% level. Elasticities are calculated at the mean values. P-values are given below each elasticity.
Source: Calculated by author.

Table 4.5. Uncompensated Elasticity Estimates for the LA/GAIDS Model

	Canned Fruit	Fresh Fruit	Frozen Fruit	Canned Vegetables	Fresh Vegetables	Frozen Vegetables	Expenditure Elasticity
Canned	**-0.7517**	**-2.7460**	**-0.9987**	0.4182	**-1.0633**	**-0.9223**	**3.0026**
Fruit	**0.0633**	**<.0001**	**<.0001**	0.1919	**0.0001**	**0.0026**	**<.0001**
Fresh	**-0.1492**	**-0.7296**	**0.1771**	**0.1581**	**0.2717**	**0.2352**	**1.9344**
Fruit	**0.0923**	**<.0001**	**0.0315**	**0.0718**	**0.0486**	**0.0100**	**<.0001**
Frozen	0.1003	**0.8765**	-0.1303	**0.8135**	0.4199	0.3298	**1.3304**
Fruit	0.7179	**0.0021**	0.6627	**0.0012**	0.1025	0.2437	**<.0001**
Canned	**0.8966**	**0.3418**	**0.3837**	**-1.4950**	**0.2543**	**0.7095**	**1.5837**
Vegetables	**<.0001**	**0.0223**	**<.0001**	**<.0001**	**0.0444**	**<.0001**	**<.0001**
Fresh	0.0731	**0.2157**	0.1000	0.0735	**-0.7605**	-0.0551	**1.6830**
Vegetables	0.5305	**0.0793**	0.3509	0.5341	**<.0001**	0.6204	**<.0001**
Frozen	**0.4593**	**0.6861**	**0.4591**	**0.8437**	0.0570	**-0.5509**	**1.4103**
Vegetables	**0.0068**	**<.0001**	**0.0004**	**<.0001**	0.6577	**0.0188**	**<.0001**

Note: The table presents estimated uncompensated elasticities from the LA/GAIDS. Elasticities in bold are significant at the 10% level. Elasticities are calculated at the mean values. P-values are given below each elasticity.
Source: Calculated by author.

Table 4.6. Compensated Elasticity Estimates for the LA/AIDS Model

	Canned Fruit	Fresh Fruit	Frozen Fruit	Canned Vegetables	Fresh Vegetables	Frozen Vegetables
Canned Fruit	-0.403 0.3357	**-1.153** **0.0003**	-0.0902 0.3106	**1.00009** **0.0003**	**0.68475** **0.0038**	-0.0388 0.9061
Fresh Fruit	**-0.1878** **0.0003**	**-0.3991** **<.0001**	0.02304 0.1638	0.02611 0.5979	**0.43721** **<.0001**	**0.10054** **0.0484**
Frozen Fruit	-0.2727 0.3106	0.42786 0.1638	-0.3119 0.2257	0.1035 0.692	0.20288 0.4357	-0.1497 0.643
Canned Vegetables	**0.43426** **0.0003**	0.06961 0.5979	0.01486 0.692	**-1.4206** **<.0001**	**0.18092** **0.0551**	**0.72095** **<.0001**
Fresh Vegetables	**0.10826** **0.0038**	**0.42449** **<.0001**	0.01061 0.4357	**0.06588** **0.0551**	**-0.5441** **<.0001**	**0.344** **<.0001**
Frozen Vegetables	-0.0164 0.9061	**0.26085** **0.0484**	-0.0209 0.643	**0.70152** **<.0001**	**-0.1742** **0.0683**	**-0.7509** **<.0001**

Note: The table presents estimated compensated elasticities from LA/AIDS and LA/GAIDS. Elasticities in bold are significant at the 10% level. Elasticities are calculated at the mean values. P-values are given below each elasticity.
Source: Calculated by author.

Table 4.7. Compensated Elasticity Estimates for the LA/GAIDS Models

	Canned Fruit	Fresh Fruit	Frozen Fruit	Canned Vegetables	Fresh Vegetables	Frozen Vegetables
Canned Fruit	-0.5891 0.1442	**-2.5834** **<.0001**	**-0.8361** **<.0001**	**0.5808** **0.0737**	**-0.9007** **0.0014**	**-0.7597** **0.0140**
Fresh Fruit	**0.4941** **0.0002**	-0.0864 0.5299	**0.8204** **<.0001**	**0.8013** **<.0001**	**0.9150** **<.0001**	**0.8785** **<.0001**
Frozen Fruit	0.1241 0.6549	**0.9003** **0.0016**	-0.1065 0.7214	**0.8373** **0.0008**	**0.4437** **0.0847**	0.3536 0.2115
Canned Vegetables	**1.0941** **<.0001**	**0.5393** **0.0006**	**0.5812** **<.0001**	**-1.2975** **<.0001**	**0.4518** **0.0008**	**0.9071** **<.0001**
Fresh Vegetables	**0.6496** **<.0001**	**0.7921** **<.0001**	**0.6765** **<.0001**	**0.6499** **<.0001**	**-0.1841** **0.0556**	**0.5213** **<.0001**
Frozen Vegetables	**0.6400** **0.0001**	**0.8669** **<.0001**	**0.6398** **<.0001**	**1.0245** **<.0001**	**0.2378** **0.0569**	-0.3702 0.1052

Note: The table presents estimated compensated elasticities from LA/AIDS and LA/GAIDS. Elasticities in bold are significant at the 10% level. Elasticities are calculated at the mean values. P-values are given below each elasticity.
Source: Calculated by author.

Uncompensated own- and cross-price demand and expenditure elasticities for the LA/AIDS are given in table 4.4. Table 4.5 presents the uncompensated own- and cross-price and expenditure elasticities for the LA/GAIDS. Compensated elasticities for the LA/AIDS are given in table 4.6. Table 4.7 provides the compensated elasticities for the LA/GAIDS. The more important result to notice from table 4.6 and table 4.7 is that 27 of 36 uncompensated elasticities for LA/GAIDS are more elastic than for LA/AIDS. Also, 29 of 36 compensated elasticities for LA/GAIDS for more elastic than for LA/AIDS. However, only one of the six compensated own-price elasticities are more elastic in LA/GAIDS than in the LA/AIDS. The two most inelastic compensated own-price elasticities are frozen fruit and fresh fruit. The own-price elasticities for fresh vegetables are quite close.

For the LA/AIDS, all uncompensated and compensated own-price elasticities are negative with those for canned fruit and frozen fruit not being significant (using a 0.10 significance level for the results section). For the LA/GAIDS, all uncompensated and compensated own-price elasticities are negative. Close to one-half of the uncompensated cross-price elasticities are significant for both of the LA/AIDS and LA/GAIDS with more LA/GAIDS elasticities being significant. Slightly more than one-half of the compensated cross-price elasticities are significant for both of the LA/AIDS. Most of the LA/GAIDS compensated elasticities are significant. Generally, the elasticities including frozen fruit are not significant with the only exception being compensated cross-price of fresh fruit. Eleven of the compensated elasticities become positive the LA/AIDS and four of the compensated elasticities for LA/GAIDS become positive.

All the expenditure elasticities presented in tables 4.5 and 4.6 are significant and positive. This implies that all the goods presented are normal (assuming the expenditure elasticity proxies for the income elasticity). For the LA/AIDS, canned fruit and fresh fruit are luxury goods while the other four are necessities. For the LA/GAIDS, the discretionary portion of all goods are considered luxury goods. All the expenditure elasticities are larger the LA/GAIDS, as we expected. This indicates that the discretionary portion for the LA/GAIDS is relatively more of a luxury good than the discretionary portion for LA/AIDS.

The uncompensated cross-price elasticities show if the good is a gross substitute (positive) or gross complement (negative) while compensated elasticities show if it is a net substitute (positive) or net complement (negative). For the LA/AIDS, the uncompensated elasticities (table 4.4) reveal that 10 are gross substitutes and 20 are gross complements. The compensated elasticities (table 4.6) reveal that 9 are net complements and 21 are gross substitutes. For the LA/GAIDS, the uncompensated elasticities (table 4.5) reveal that 24 are gross substitutes and 6 are gross complements. The compensated elasticities (table 4.7) reveal that 5 are net complements and 25 are gross substitutes.

For both models, canned fruit is a gross complement with fresh fruit, frozen fruit, fresh vegetables, and frozen vegetables. Fresh fruit and canned fruit are gross complements for both models. Fresh vegetables and frozen vegetables are gross complements. Canned fruit is a gross substitute with canned vegetables. Frozen fruit is a gross substitute with all other types of products. Fresh fruit and fresh vegetables are gross substitutes. For the LA/AIDS, fresh fruit and frozen vegetables are gross complements while they are gross substitutes for the LA/GAIDS.

In terms of net substitutes and complements, the results within produce type are similar for both models. For both models, canned fruit is a net complement with fresh fruit and frozen fruit. Fresh fruit and frozen fruit are net substitutes. Canned vegetables are net substitutes with fresh vegetables and frozen vegetables. The results for fresh vegetables indicate that it is a net substitute with all other types of products. Canned fruit is a net substitute for canned vegetables for both models. For both models, canned fruit and frozen vegetables are net complements. Fresh fruit is a net complement for canned fruit in the LA/AIDS, but is a net substitute for LA/GAIDS. Frozen fruit and frozen vegetables are net complements for LA/AIDS and net substitutes for LA/GAIDS.

It is informative to compare our results with elasticity estimates for fruit and vegetables from other authors. Our estimates for our LA/AIDS are similar to those in the literature. Park et al. (1996) find own price elasticities of -0.34 for fruit and -0.32 for vegetables for low-income households. Dong and Lin (2009) find own-price elasticities of -0.52 for fruit and -0.69 for vegetables for low-income households. Our estimates of -0.76 and -0.73 for fresh fruit are slightly higher but still on the inelastic side in a similar way to these authors. Our own price elasticity measurements for fresh vegetables also follow a similar pattern as those of fresh fruit.

Using the uncompensated elasticities (table 4.4 and table 4.5), the effect of a policy on fruit and vegetable purchases can be calculated. We set the baseline purchases (grams per month) as the average of last 12 months of the per capita purchases in each respective category. For the pre-committed amount, we also used the average level of pre-commitment over the last 12 months for each category. A negative level of pre-commitment was present in canned fruit so this was set to zero for the policy analysis and represents no pre-commitments in canned fruit.[13] The full baseline purchases are able to be impacted by price changes for the LA/AIDS and are the discretionary portions. For the LA/GAIDS, the discretionary portions are the baseline purchases minus the estimated level of pre-commitments.

Table 4.8 shows the effects of a subsidy that would result in a 20% reduction in price faced by consumers applied all categories of fruit and vegetable products. The subsidy predicts increases in purchases across all categories for both models. The increase is less in the LA/GAIDS model since the discretionary portion of purchases is smaller than for LA/AIDS. For example, the discretionary portion for fresh fruit in the LA/AIDS calculation is 1590.2 grams per month while the discretionary portion for LA/GAIDS is 879.3 grams per month. Even though the elasticities are relatively more elastic in the LA/GAIDS, the change takes place on a smaller baseline and leads to a smaller predicted increase in purchases. The bottom rows show this difference quite clearly. For example, the subsidy leads to a smaller predicted increase in the purchase of fresh fruit by 337.7 grams per month. The LA/AIDS gives higher predicted increases in purchases ranging from 11.3% to 29.0%. LA/AIDS predicted a higher purchase of fresh

[13] Recall that Tonsor and Marsh (2007) explain negative estimates of pre-committed demand as a preference to shift out of pre-committed expenditure and into supernumerary expenditures (those that are not affected by price). Thus, a negative value for pre-commitments indicates the consumer is more affected by economic factors, such as price. Setting the level of pre-commitment to zero for canned fruit means that the entire baseline level is now discretionary.

fruit by 17.5% and a higher purchase of fresh vegetables by 11.3% relative to the LA/GAIDS. Overall, the LA/AIDS predicts a higher total increase in the purchase of fruit and vegetable consumption by 709.9 grams per month than LA/GAIDS with the same subsidy.

Table 4.8. Change in Grams/Month Consumed from a Proposed Subsidy Resulting in a 20% Drop in Price to all Products

	Canned Fruit	Fresh Fruit	Frozen Fruit	Canned Vegetables	Fresh Vegetables	Frozen Vegetables
LA/AIDS						
Discretionary level	172.6	1590.2	45.2	578.0	1492.2	388.8
Change in discretionary	85.4	344.2	7.6	100.9	288.4	31.2
New purchase level	258.0	1934.4	52.8	678.9	1780.6	420.0
LA/GAIDS						
Pre-committed level	0.0	710.9	29.2	242.9	249.3	196.8
Discretionary level	172.6	879.3	16.0	335.1	1242.9	192.0
Change in discretionary	121.3	6.5	-7.7	-73.1	87.8	-75.1
New purchase level	381.9	1596.7	37.5	504.9	1580.0	313.8
Difference in predicted purchase	-123.9	337.7	15.3	174.0	200.6	106.2
Percentage difference	-48.0	17.5	29.0	25.6	11.3	25.3
Total difference = 709.9						

Note: This scenario is a subsidy that would result in a 20% reduction in price applied to all fruit and vegetables. Quantities are in grams purchased per month per capita. The discretionary level of grams per day is the average of last 12 months of the per capita purchase in the respective category. This discretionary baseline amount is increased or decreased by the corresponding own and cross price elasticities to find the percent change for each category. The difference is found by subtracting the predicted amount of LA/GAIDS from the predicted amount from LA/AIDS. The percentage difference is the percent higher/lower that LA/AIDS predicts over LA/GAIDS.
Source: Calculated by author.

Table 4.9 shows the effects of a subsidy resulting in a 20% drop in price applied to only fresh fruit and fresh vegetables. These were chosen since they are products that policy makers want citizens to increase in consumption. The subsidy predicts increases in purchases for both models across all products except frozen fruit (likely due to the many statically insignificant elasticities for this product). Similar to table 4.8, the LA/AIDS predicts are larger increase in purchases for many of the products due to the higher discretionary baseline. One of the exceptions to this is canned fruit. LA/GAIDS predicts that consumers' purchases of canned fruit will be 57.7 grams higher per month than forecast by LA/AIDS. For the other five categories, LA/GAIDS predicts lower purchases relative to the LA/AIDS.

For the fresh fruits and fresh vegetables, the products directly affected by the subsidy, LA/AIDS predicts that consumers will purchase 220.0 and 230.4 grams per month more respectively. This compares to the 80.5 and 135.4 grams per month predicted increase from the LA/GAIDS. Thus, LA/AIDS predicts a higher purchase of fresh fruit by 7.7% and a higher purchase of fresh vegetables by 10.6% relative to the LA/GAIDS. Overall, the LA/AIDS predicts that consumers will increase their purchases of fruits and vegetables by 302.5 grams per month than the prediction of LA/GAIDS with the same subsidy. Given our situation and assumptions, ignoring the pre-commitments would mislead policy makers into believing a predicted policy would lead to a larger increase in the purchase of fruits and vegetables relative to when pre-commitments are included.

Table 4.9. Change in Grams/Month Consumed from a Proposed Subsidy Resulting in a 20% Drop in Price to Fresh Products

	Canned Fruit	Fresh Fruit	Frozen Fruit	Canned Vegetables	Fresh Vegetables	Frozen Vegetables
LA/AIDS						
Discretionary level	172.6	1590.2	45.2	578.0	1492.2	388.8
Change in discretionary	73.8	220.0	-0.6	39.1	230.4	14.3
New purchase level	246.4	1810.4	44.6	617.1	1722.6	403.1
LA/GAIDS						
Pre-committed level	0.0	710.9	29.2	242.9	249.3	196.8
Discretionary level	172.6	879.3	16.0	335.1	1242.9	192.0
Change in discretionary	131.5	80.5	-4.2	-40.0	135.4	-28.5
New purchase level	304.1	1670.7	41.0	538.0	1627.6	360.3
Difference in predicted purchase	-57.7	139.7	3.6	79.1	95.0	42.8
Percentage difference	-23.4	7.7	8.1	12.8	5.5	10.6
Total difference = 302.5						

Note: This scenario is subsidy that would result in a 20% reduction in price applied to only fresh fruit and vegetables. Quantities are in grams purchased per month per capita. The discretionary level of grams per day is the average of last 12 months of the per capita purchase in the respective category. This discretionary baseline amount is increased or decreased by the corresponding own and cross price elasticities to find the percent change for each category. The difference is found by subtracting the predicted amount of LA/GAIDS from the predicted amount from LA/AIDS. The percentage difference is the percent higher/lower that LA/AIDS predicts over LA/GAIDS.
Source: Calculated by author.

The results of the proposed subsidy may initially be confusing given that more elasticities for LA/GAIDS are more elastic than for LA/AIDS. However, it needs to be re-emphasized that even though the elasticities are relatively more elastic in the LA/GAIDS, the changes from the subsidy take place on only on the discretionary portion of purchases. These discretionary portions are smaller for the LA/AIDS and thus these larger elasticity estimates are offset by the higher levels of pre-commitments found in many of our products. Had our estimates of pre-commitments been small, it is likely that the LA/GAIDS would predict higher increases in purchases. In addition, it is important to note that we only used the average level of pre-commitment for the previous 12 months. Figure 4.4 shows that the pre-commitments can vary over time and are even equal to zero in some months. During the months with zero pre-commitments, the LA/GAIDS would certainly predict higher levels of purchases than the LA/AIDS.

Conclusions and Policy Implications

The results of this study indicate a need to account for pre-committed demand when examining consumer food demand. To estimate the size of policy effects it is necessary to specify the correct functional form. To demonstrate this we compare and AIDS with a GAIDS, which incorporates pre-commitments. The incorporation of pre-commitments (non-discretionary) allows the separation of the portion of consumption that is not affected by changes in price and the discretionary level. The discretionary level (total consumption minus the pre-committed amount) is the portion of total consumption that is affected by changes in price. For AIDS since there is no pre-commitment, then the total consumption for a commodity is the same as the discretionary consumption for that commodity. As expected, a major result from this study is that elasticities calculated under the presence of pre-commitments (GAIDS) are more elastic relative to those calculated without.

The results from a proposed subsidy further reinforce the importance of accounting for pre-commitments. For a 20% reduction in price due to a subsidy applied to all products, LA/AIDS predicts a higher increase in the purchase of fresh fruit by

17.5% and a higher increase in the purchase of fresh vegetables by 25.3% relative to the LA/GAIDS. Even though the elasticities are more elastic, the increases in purchases are smaller for the LA/GAIDS than the LA/AIDS since the discretionary portion of purchases is smaller for LA/GAIDS. The LA/AIDS gives higher predicted increases in purchases than LA/GAIDS ranging from 11.3% to 29.0% higher. When comparing the forecasted increase in purchases for both models with the same subsidy, the LA/AIDS predicts a larger increase in the purchase of fruits and vegetables by 709.9 grams per month than the LA/GAIDS.

Similar results hold for a 20% reduction in price due to a subsidy applied to only fresh fruit and fresh vegetables. For the fresh fruit and fresh vegetables, LA/AIDS predicts that consumers will purchase 220.0 and 230.4 grams per month more respectively. This compares to the 80.5 and 135.4 grams per month predicted increase from the LA/GAIDS. Thus, LA/AIDS predicts 7.7% more purchases of fresh fruit 5.5% more purchases of fresh vegetables relative to the LA/GAIDS. Overall, the LA/AIDS predicts that consumers will increase their purchases of fruits and vegetables by 302.5 grams per month more than the prediction of LA/GAIDS with the same subsidy. Failure to include pre-commitments would cause a predicted policy to a forecast a larger increase in the purchase of fruits and vegetables relative to when pre-commitments are included.

One further important result from this policy analysis is not only the need to account for pre-commitments, but also the need to account for the consumer's timing of pre-commitments. Figure 4.4 illustrates how the level of pre-commitments vary throughout the year. A policy that affects the price of fruit or vegetables products will have the greatest effect during months with low levels of pre-commitments. A policy that affects price during months with high pre-commitment will have a small impact and will be least likely to influence consumer behavior. The policy analysis in this study used the average of the last 12 months of commitments. Further research could be done to account for month-to-month change in the level of pre-commitment. This would allow the comparison of outcomes for a flat subsidy every month versus one that increases in months with lower pre-commitments.

In terms of compliance with theoretical regularity conditions, the GAIDS performs slightly better than the AIDS. Positivity is satisfied at every point for the LA/AIDS and the LA/GAIDS. Monotonicity is violated at six points for the LA/AIDS and at zero points for the LA/GAIDS. The curvature condition was violated at every point for both models. An attempt was made to apply the curvature conditions using the method of Ryan and Wales (1998). The statistical program used for this was not able to find a closed form solution. The extra complexity this method imposes likely caused issues with the numerical estimation procedure.

Some limitations remain with this study. Given the inherent limitation attributed to Nielsen Homescan Consumer Panel data, the focus of this paper is food purchased for consumption at home. Produce consumed away from home would not be captured by this dataset. This may not be a major problem as eating meals away from home is usually associated with eating less produce products (Lin and Guthrie, 2012; Todd, Mancino and Lin, 2010) and this might not change overall produce totals. This dataset does not provide time spent preparing food and only includes food purchases. The need to account for time is especially important since food prices influence food production and time allocation decisions (Aguiar and Hurst, 2007; Senia, Jensen, and Zhylyevskyy, 2017). One must be careful to differentiation between food that is purchased and food that is consumed.

Results from this research do raise a few lines of further research in the area of pre-commitments. Our estimation estimated the presence of negative pre-commitment values. The presence and implications of these estimated negative pre-commitments in an area in need of further research. Further, this research would benefit from the addition of curvature so that both models satisfy this regularity condition. Given the results of our policy analysis, more work should be done on the timing of pre-commitments. This research could explore how the timing of pre-commitments could affect the results of a policy analysis.

CHAPTER V

CONCLUSIONS

The overall objective of this dissertation was to contribute to a better understanding of consumer food acquisitions by: (1) considering policies to promote dietary fiber consumptions; (2) modeling consumer food acquisitions as a complex economic system; and (3) determining the effects of pre-determined demand and theoretical regularity conditions of models on policy analysis. To achieve these objectives, three related empirical investigations of consumer food acquisitions were conducted (Chapters II-IV).

This first paper contributed to the literature by conducting a panel regression on nine per-capita fiber intake categories taken from purchases of variety of food types such as bread, pasta, tortilla, fresh fruit, fresh vegetables and beans, frozen fruit, frozen vegetables and beans, canned fruit, and canned vegetables and beans to uncover socioeconomic and government food policy related factors on the per capita intake of dietary fiber in the United States. A number of interesting finding resulted from the analysis. For the mean household in the sample there is a 5.5% decrease in per capita dietary fiber purchased in the period after the dietary guidelines were released. A proposed 20% subsidy applied to fruits and vegetables would result in an increase in the average per capita consumption of fiber per day by 4.8%. Although consumer response to 2010 *Dietary Guidelines for Americans* in terms of increased intake of dietary fiber showed mixed results, a proposed 20% subsidy on fruits and vegetables showed some promising results concerning increasing fiber intake.

The contribution of the second was to use the individual and household characteristics, characteristics of the local food environment, the individual's dietary pattern, prices, health outcomes, and policy variables jointly to estimate a complex graphical causality structure. In regards to the paths between poverty, race and food insecurity, we find a number of paths. We find that Hispanic individuals are more likely to be food insecure. There is also a direct path between the percent of poverty level and food insecurity and a path between college education and food insecurity. We find a causal chain from Black to SNAP participation to low food security. A similar casual chain also exists for Hispanic individuals to low food security via SNAP participation. These results mean that if SNAP participation is included in a model with these race variables and food insecurity, the path between race and food insecurity will be blocked by the inclusion of SNAP participation. Thus, policymakers that want to reduce the problems associated with obesity or food insecurity need a full picture of the complex interactions among all these variables.

The objective of the third paper was to examine the affect that ignoring pre-determined demand and theoretical regularity conditions will have on consumer food demand. To accomplish this we used the AIDS because of its wide use in applied policy research. A major result from this study is that elasticities calculated under the presence of pre-commitments are more elastic relative to those calculated without. However, applying these elasticities in a proposed policy may not lead to a larger predicted change in quantity if there are high levels of pre-commitments because the policy will only affect the discretionary portion of purchases. The result from a proposed subsidy further reinforces the importance of accounting for pre-commitments. For a subsidy that resulted in a 20% price reduction applied to all products, the AIDS with no pre-commitments predicts 292.8 grams per month larger increase in the purchases of fruits and vegetables than the prediction from the AIDS with pre-commitments. For a the subsidy applied to only fresh fruit and vegetables, the AIDS with no pre-commitments predicts 165.2 grams per month more fruits and vegetables will be purchased than the AIDS with pre-commitments. In terms of satisfying regularity conditions, both models satisfy positivity, the AIDS with pre-commitments performs slightly better in terms of monotonicity, and both models fail to satisfy local curvature. One further important result from this study is not only the need to account for pre-commitments, but also the need to account for the consumer's timing of pre-commitments.

REFERENCES

Abdulai, A., and D. Aubert. 2004. “A Cross-Section Analysis of Household Demand for Food and Nutrients in Tanzania.” *Agricultural Economics* 31(1):67-79.

Aguiar, M., and E. Hurst. 2005. “Consumption versus Expenditure.” *Journal of Political Economy* 113(5):919-948.

Ahima, R.S., and M.A. Lazar. 2013. “The Health Risk of Obesity-Better Metrics Imperative.” *Science* 23:856-858.

Alston, J. M., J.A. Chalfant, and N.E. Piggott. 2001. “Incorporating Demand Shifters in the Almost Ideal Demand System.” *Economics Letters* 70:73-78.

Alston, J.M., K.A. Foster, and R.D. Green. 1994. “Estimating Elasticities with the Linear Approximate Almost Ideal Demand System: Some Monte Carlo Results.” *Review of Economics and Statistics* 76(2):351-356.

Alviola, P.A., and O. Capps Jr. 2010. “Household Demand Analysis of Organic and Conventional Fluid Milk in the United States Based on the 2004 Nielsen Homescan Panel.” *Agribusiness* 26(3):369-388.

Amemiya, T. 1984. “Tobit Models: A Survey.” *Journal of Econometrics* 24(1-2):3-61.

Anderson, J.W., P. Baird, R.H. Davis, S. Ferreri, M. Knudtson, A. Koraym, V. Waters, and C.L Williams. 2009. “Health Benefits of Dietary Fiber.” *Nutrition* 67(4):188-205.

Anderson, T.W., and D.A. Darling. 1952. "Asymptotic Theory of Certain "Goodness-of-Fit" Criteria Based on Stochastic Processes." *Annals of Mathematical Statistics* 23:193–212.

Andrews, M., R. Bhatta, and M.V. Ploeg. 2013. “An Alternative to Developing Stores in Food Deserts: Can Changes in SNAP Benefits Make a Difference?” *Applied Economic Perspectives and Policy* 35(1):150-170.

Andreyeva, T., and J. Luedicke. 2013. “Federal Food Package Revisions: Effects of Purchases of Whole-Grain Products.” *American Journal of Preventative Medicine* 45(4):422-429.

Banks, J., R. Blundell, and A. Lewbel. 1997. “Quadratic Engel Curves and Consumer Demand.” *Review of Economics and Statistics* 79(4):527-539.

Barlow, S.E. 2007. “Expert Committee Recommendations Regarding the Prevention, Assessment, and Treatment of Child and Adolescent Overweight and Obesity: Summary Report.” *Pediatrics* 120(Supplement 4):S164-S192.

Barnett, W.A. 2002. “Tastes and Technology: Curvature is not Sufficient for Regularity.” *Journal of Econometrics* 108: 199-202.

Barnett, W.A., and A. Serletis. 2008. “Consumer Preferences and Demand Systems.” *Journal of Econometrics* 147(2):210-224.

Basu, S., C. Wimer, and H. Seligman. 2016. “Moderation of the Relation of County-Level Cost of Living to Nutrition by the Supplemental Nutrition Assistance Program.” *American Journal of Public Health* 106(11):2064-2070.

Beatty, T.K.M., B.H. Lin, and T.A. Smith. 2014. “Is Diet Quality Improving? Distributional Changes in the United States, 1989-2008.” *American Journal of Agricultural Economics* 96(3):769-789.

Berndt, E. R., and N. E. Savin. 1975. "Estimation and Hypothesis Testing in Singular Equation Systems with Autoregressive Disturbances." *Econometrica* 43(5/6):937-57.

Bessler, D.A., and S. Lee. 2002. “Money and Prices: U.S. Data 1869-1914 (A Study with Directed Graphs).” *Empirical Economics* 27:427-446.

Beydoun, M.A., L.M. Powell, X. Chen, and Y. Wang. 2011. “Food Prices Are Associated with Dietary Quality, Fast Food Consumption, and Body Mass Index among U.S. Children and Adolescents.” *The Journal of Nutrition* 141:304–311.

Bollen, K. A. 1989. *Structural Equations with Latent Variables.* New York: John Wiley & Sons.

Bollino, C.A. 1987. “GAIDS: A Generalized Version of the Almost Ideal Demand System.” *Economics Letters* 23:199-202.

Bradlee, M.L., M.R. Singer, M.M. Quresh, and L.L. Moore. 2009. “Food Group Intake and Central Obesity among Children and Adolescents in the Third National Health and Nutrition Examination Survey (NHANES III).” *Public Health Nutrition* 13(6):797–805.

Capps Jr., O., R. Tsai, R. Kirby, and G. Williams. 1994. “A Comparison of Demand for Meat Products in the Pacific Rim Region.” *Journal of Agricultural and Resource Economics* 19(1):210-224.

Chang, D., and A. Serletis. 2012. “Imposing Local Curvature in the QUAIDS.” *Economics Letters* 115:41-43.

Chang, D., and A. Serletis. 2013. “The Demand for Gasoline: Evidence from Household Survey Data.” *Applied Econometrics* 29(2):291-313.

Chen, D., E.C. Jaenicke, and R.J. Volpe. 2016. “Food Environments and Obesity: Household Diet Expenditure versus Food Deserts.” *American Journal of Public Health* 106(5):881-888.

Chickering, D.M. 2002. “Optimal Structure Identification with Greedy Search.” *Journal of Machine Learning Research* 3:507-554.

Chickering, D.M., and C. Meek. 2002. “Finding Optimal Bayesian Networks.” In *Proceeding of the Eighteenth Conference on Uncertainty in Artificial Intelligence.* Morgan Kaufmann Publishers, pp. 94-102.

Christensen, L. R., Jorgenson, D. W., Lau, L. J. 1973 “Transcendental Logarithmic Productions Frontiers.” *Review of Economics and Statistics* 55:28-45.

Cox, T., and M. Wohlgenant. 1986. “Prices and Quality Effects in Cross-Section Demand Analysis.” *American Journal of Agricultural Economics* 68:908-919.

Cutler, D.M., A. Lleras-Muney, and T. Vogl. 2011. “Socioeconomic Status and Health: Dimensions and Mechanisms.” In S. Glied and P.C. Smith, ed. *Oxford Handbook of Health Economics*. Oxford, UK: Oxford University Press, pp. 124-168.

Darmon, N. and A. Drewnowski. 2008. “Does Social Class Predict Diet Quality?” *American Journal of Clinical Nutrition* 87:1107-1117.

Davis, G.A. 2003. “Bayesian Reconstruction of Traffic Accidents.” *Law, Probability and Risk* 2:69-89.

Deaton, A., and J. Muellbauer. 1980a. “An Almost Ideal Demand System.” *American Economic Review* 70:312-326.

Deaton, A., and J. Muellbauer. 1980b. *Economics and Consumer Behavior*. Cambridge University Press.

Deaton, A. 2016. “Measuring and Understanding Behavior, Welfare, and Poverty.” *American Economic Review* 106(6):1221-1243.

Dharmasena, S., and O. Capps Jr. 2012. "Intended and Unintended Consequences of a Proposed National Tax on Sugar-Sweetened Beverages to Combat the U.S. Obesity Problem." *Health Economics* 21(6):669-694.

Dharmasena, S. and O. Capps Jr. 2014. "Unraveling Demand for Dairy-Alternative Beverages in the United States: The Case of Soymilk." *Agricultural and Resource Economics Review* 43(1):140-157

Dharmasena, S., O. Capps Jr., and A. Clauson. 2011. "Ascertaining the Impact of 2000 USDA Dietary Guidelines for Americans on the Intake of Calories, Caffeine, Calcium and Vitamin C from At-Home Consumption of Nonalcoholic Beverages." *Journal of Agricultural and Applied Economics* 43(1):13-27

Dharmasena, S., D.A. Bessler, and O. Capps Jr. 2016. "Food Environment in the United States as a Complex Economic System." *Food Policy* 61:163-175.

Diewert, W.E., and T.J. Wales. 1988. "Normalized Quadratic Systems of Consumer Demand functions." *Journal of Business and Economic Statistics* 6:303-312.

Dong, D., and B.H. Lin. 2009. *Fruit and Vegetable Consumption by Low-Income Americans: Would a Price Reduction Make a Difference?* Washington DC: Economic Research Service, U.S. Department of Agriculture, Economic Research Report Number 70, January.

Dubowitz, T., M. Heron, C.E. Bird, N. Lurie, B.K. Finch, R. Basurto-Da'vila, L. Hale, and J.J. Escarce. 2008. "Neighborhood Socioeconomic Status and Fruit and Vegetable Intake among Whites, Blacks, and Mexican Americans in the United States." *American Journal of Clinical Nutrition* 87:1883-1891.

Einav, L., E. Leibtag, and A. Nevo. 2010. "Recording Discrepancies in Nielsen Homescan Data: Are they Present and do They Matter?" *Quantitative Marketing in Economics* 8(2):207-239.

Epstein, L.H., K.K. Dearing, L.G. Roba, and E. Findelstein. 2010. "The Influence of Taxes and Subsidies on Energy Purchased in an Experimental Purchasing Study." *Psychological Science* 21:406-414.

Finkelstein, E.A., C.J. Ruhm, and K.M. Kosa. 2005. "Economic Causes and Consequences of Obesity." *Annual Review of Public Health* 26:239-257.

Fung, T.T., F.B. Hu, M.A. Pereira, S. Liu, M.J. Stampfer, G.A. Colditz, and W.C. Willett. 2002. "Whole-Grain Intake and the Risk of Type 2 Diabetes: A Prospective Study in Men." *American Journal of Clinical Nutrition* 76:535-540.

Greene, W. H. 2012. *Econometric Analysis*, 7th ed. Upper Saddle River, NJ: Prentice Hall.

Garasky, S., L.W. Morton, and K.A. Greder. 2006. "The Effects of the Local Food Environment and Social Support on Rural Food Insecurity." *Journal of Hunger and Environmental Nutrition* 1(1):83-103.

Geiger, D., T. Verma, and J. Pearl. 1990. "Identifying Independencies in Bayesian Networks." *Networks* 20:507-534.

Glymour, C., R. Scheines, P. Spirtes, and J. Ramsey. 2016. "The TETRAD Project: Causal Models and Statistical Data." Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA, http://www.phil.cmu.edu/tetrad/current.html.

Gould, B.W., and H.J. Villarreal. 2002. "Adult Equivalence Scales and Food Expenditures: An Application to Mexican Beef and Pork Purchases." *Applied Economics* 34:1075-1108.

Gray, K.F. 2014. *Characteristics of Supplemental Nutrition Assistance Program Households: Fiscal Year 2013.* Washington DC: U.S. Department of Agriculture, Food and Nutrition Service, Supplemental Nutrition Assistance Program Report No. SNAP-14-CHAR, December.

Gundersen, C., and V. Oliveira. 2001. "The Food Stamp Program and Food Insufficiency." *American Journal of Agricultural Economics* 83(4):875-887.

Gundersen, C., and J.P. Ziliak. 2015. "Food Insecurity and Health Outcomes." *Health Affairs* 34(11):1830-1839.

Gundersen, C., B. Kreider, and J. Pepper. 2011. "The Economics of Food Security in the United States." *Applied Economic Perspectives and Policy* 33(3):281-303.

Gundersen, C., A. Dewey, A.S. Crumbaugh, M. Kato, and E. Engelhard. 2014. *Map the Meal Gap 2016: Technical Brief*, Feeding America.

Guthrie, J.F., B.H. Lin, and E. Frazao. 2002. "Role of Food Prepared Away From Home in the American Diet, 1977-78 versus 1994-96: Changes and Consequences." *Journal of Nutrition Education and Behavior* 34(3):140-150.

U.S. Department of Health & Human Services (HHS). 2012. "2012 HHS Poverty Guidelines." *Federal Register* (77)17:4034-4035

Hovhannisyan, V., and B.W. Gould. 2011. "Quantifying the Structure of Food Demand in China: An Econometric Approach." *Agricultural Economics* 42:1-17.

Hovhannisyan, V., and B.W. Gould. 2014. “Structural Change in Urban Chinese Food Preferences.” *Agricultural Economics* 45:159-166.

Hoy, M.K., and J.D. Goldman. 2010. *Fiber Intake of the U.S. population: What We Eat in America, NHANES 2009-2010.* Washington DC: Food Surveys Research Group, U.S. Department of Agriculture, Data Brief No. 12, September.

Huang, H.C. 2001. “Bayesian Analysis of the SUR Tobit Model.” *Applied Economics Letters* 8(9):617-622.

Jensen, H.H. 2002. “Food Insecurity and the Food Stamp Program.” *American Journal of Agricultural Economics* 84(5):1215-1228.

Kennedy, P. 2003. Limited Dependent Variables. *A Guide to Econometrics*, MIT Press.

Kilts Center for Marketing. 2013. *Consumer Panel Dataset Manual*, Booth School of Business, University of Chicago.

Kim, S.Y., R.M. Nayga Jr., and O. Capps Jr. 2000. “The Effect of Food Label Use on Nutrient Intakes: An Endogenous Switching Regression Analysis.” *Journal of Agricultural and Resource Economics* 25(1):215-231.

King, D.E., A.G. Mainous, and C.A Lambourne. 2012. “Trends in Dietary Fiber Intake in the United States.” *Journal of the Academy of Nutrition and Dietetics* 112(5):642-648.

Klerman, J.A., S. Bartlett, P. Wilde, and L. Olsho. 2014. “The Short-Run Impact of the Health Incentives Pilot Program on Fruit and Vegetable Intake.” *American Journal of Agricultural Economics* 96(5):1372-1382.

Krishnan, S., L. Rosenberg, M. Singer, F.B. Hu, L. Djoussé, L.A. Cupples, and J.R. Palmer. 2007. “Glycemic Index, Glycemic Load, and Cereal Fiber Intake and Risk of Type 2 Diabetes in U.S. Black Women.” *Archives of Internal Medicine* 167:2304-2309.

Kwate, N.O.A. 2008. “Fried Chicken and Fresh Apples: Racial Segregation as a Fundamental Cause of Fast Food Density in Black Neighborhoods.” *Health and Place* 14:32-44.

Kwon, D.H., and D.A. Bessler. 2011. “Graphical Methods, Inductive Causal Inference, and Econometrics: A Literature Review.” *Computational Economics* 38(1):85-106.

Kyureghian, G., O. Capps Jr, and R.M. Nayga Jr. 2011. “A Missing Variable Imputation Methodology with an Empirical Application.” *Advances in Econometrics* 27(1):313-337.

Lai, P., and D.A. Bessler. 2015. “Price Discovery between Carbonated Soft Drink Manufacturers and Retailers: A Disaggregate Analysis with PC and LiNGAM Algorithms.” *Journal of Applied Economics* 18(1):173-197.

Lin, B.H., and J. Guthrie. 2012. *Nutritional Quality of Food Prepared at Home and Away from Home, 1977-2008.* Washington DC: Economic Research Service, U.S. Department of Agriculture, Economic Information Bulletin 105, 22 pp. December.

Lin, B.H., and S.T. Yen. 2008. “Consumer Knowledge, Food Label use and Grain Consumption in the US.” *Applied Economics* 40(4):437-448.

Lin, B.H., T.A. Smith, J.Y. Lee, and K.D. Hall. 2011. “Measuring Weight Outcomes for Obesity Intervention Strategies: The Case of a Sugar-Sweetened Beverage Tax.” *Economics and Human Biology* 9:329-341.

Lin, B.H., S.T. Yen, D. Dong, and A.M. Smallwood. 2010. “Economic Incentives for Dietary Improvement among Food Stamp Recipients.” *Contemporary Economic Policy* 28:524-536.

Liu, J.L., B. Han, and D.A. Cohen. 2015. “Beyond Neighborhood Food Environments: Distance Traveled to Food Establishments in 5 US Cities, 2009–2011.” *Preventing Chronic Disease* 12(6):E126

Liu, M., P. Kasteridis, and S.T. Yen. 2013. “Breakfast, Lunch, and Dinner Expenditures Away From Home in the United States.” *Food Policy* 38:156-164.

Liu, S., W.C. Willett, M.J. Stampfer, F.B. Hu, L. Sampson, C.H. Hennekens, and J.E. Manson. 2000. “A Prospective Study of Dietary Glycemic Load, Carbohydrate Intake, and Risk of Coronary Heart Disease in U.S. women.” *American Journal of Clinical Nutrition* 71:1455-146.

Lucan, S.C., and N. Mitra. 2012. “Perceptions of the Food Environment are Associated with Fast-Food (Not Fruit-And-Vegetable) Consumption: Findings from Multilevel Models.” *International Journal of Public Health* 57:599–608.

Maddala, G.S. 1987. “Limited Dependent Variable Models using Panel Data.” *Journal of Human Resources* 22(2):207-338.

Mancino, L., and F. Kuchler. 2012. "Demand for Whole-Grain Bread before and after the Release of the Dietary Guidelines." *Applied Economic Perspectives and Policy* 34(1):76-101

Mancino, L., F. Kuchler, and E. Leibtag. 2008. "Getting Consumers to Eat More Whole-Grains: The Role of Policy, Information, and Manufacturers." *Food Policy* 33(6):489–496.

Michels, K.B., B.R. Bloom, P. Riccardi, B.A. Rosner, and W.C. Willett. 2008. "A Study of the Importance of Education and Cost Incentives on Individual Food Choices at the Harvard School of Public Health Cafeteria." *Journal of the American College of Nutrition* 27(1):6-11.

Miguel, H.O., and D. Diansheng. 2012. "Assessing U.S. Household Purchase Dynamics for Dietary Fiber." Paper presented at AAEA annual meeting, Seattle, Washington, August 12-14.

McKeown, N.M., J.B. Meigs, S. Liu, P.W. Wilson, and P.F. Jacques. 2002. "Whole-Grain Intake is Favorably Associated with Metabolic Risk Factors for Type 2 Diabetes and Cardiovascular Disease in the Framingham Offspring Study." *American Journal of Clinical Nutrition* 76(2):390-398.

Meek, C. 1995. "Causal Inference and Causal Explanation with Background Knowledge." *In Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI-95)*, Morgan Kaufmann Publishers, Inc. pp. 403-411.

Meek, C. 1997. "Graphical Structures: Selecting Causal and Statistical Models." PhD thesis, Carnegie Mellon University.

Mehta, N.K., and V.W. Chang. 2008. "Weight Status and Restaurant Availability: A Multilevel Analysis." *American Journal of Preventative Medicine* 34(2):127-133.

Moore, L.V., A.V.D. Roux, J.A. Nettleton, and D.R. Jacobs. 2008. "Associations of the Local Food Environment with Diet Quality—A Comparison of Assessments based on Surveys and Geographic Information Systems: The Multi-Ethnic Study of Atherosclerosis." *American Journal of Epidemiology* 167(8):917-924.

Morland, K., S. Wing, and A.D. Roux. 2002. "The Contextual Effect of the Local Food Environment on Residents' Diets: The Atherosclerosis Risk in Communities Study." *American Journal of Public Health* 92(11):1761-1767.

Mumford, J.A. and J.D. Ramsey. 2014. "Bayesian Networks for fMRI: A Primer." *NeuroImage* 86:571-582.

Nevo, A. 2000. “A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand.” *Journal of Economics and Management Strategy* 9(4):513-548.

Nnoaham, K.E., G. Sacks, M. Rayner, O. Mytton, and A. Gray. 2009. “Modelling Income Group Differences in the Health and Economic Impacts of Targeted Food Taxes and Subsidies.” *International Journal of Epidemiology* 38(5):1324-1333.

Nord, M., A. Coleman-Jensen, M. Andrews, and S. Carlson, 2010. *Household Food Security in the United States, 2009*. Washington DC: U.S. Department of Agriculture, Economic Research Service, Economic Research Report No. 108.

Nordström J., and L. Thunström. 2011. “Can Targeted Food Taxes and Subsidies Improve the Diet? Distributional Effects among Income Groups.” *Food Policy* 36:259–271.

Ogden C.L., Carroll M.D., Kit B.K., and K.M. Flegal. 2014. “Prevalence of Childhood and Adult Obesity in the United States, 2011-2012.” *Journal of the American Medical Association* 311(8):806-814.

Ollberding, N.J., R.L. Wolf and I. Contento. 2011. “Food Label use and its Relation to Dietary Intake among U.S. Adults.” *Journal of the American Dietetic Association* 110(8):1223-1237.

Palma, M.A., and K.M. Jetter. 2012. “Will the 2010 Dietary Guidelines for Americans be any more effective for Consumers?” *Choices* 27(1). Available online: http://choicesmagazine.org/choices-magazine/theme-articles/potential-impacts-of-2010-dietary-guidelines-for-americans-/will-the-2010-dietary-guidelines-for-americans-be-any-more-effective-for-consumers

Park, J., R. Holcomb, K. Raper, and O. Capps Jr. 1996. “A Demand System Analysis of Food Commodities by U.S. Households Segmented by Income.” *American Journal of Agricultural Economics* 78(2):290-300.

Pearl, J. 1986. “Fusion, Propagation, and Structuring in Belief Networks” *Artificial Intelligence* 29:241-288.

Pearl, J. 1995. “Causal Diagrams for Empirical Research.” *Biometrika* 82(4):668-710.

Pearl, J. 2009. *Causality: Models, Reasoning, and Inference*, 2nd ed. New York: Cambridge University Press.

Pereira, M.A., E. O'Reilly, K. Augustsson, G.E. Fraser, U. Goldbourt, B.L. Heitmann, G. Hallmans, P. Knekt, S. Liu, P. Pietinen, D. Spiegelman, J. Stevens, J. Virtamo, W.C. Willett, and A. Ascherio. 2004. "Dietary Fiber and Risk of Coronary Heart Disease: A Pooled Analysis of Cohort Studies." *Archives of Internal Medicine* 164(4):370-376.

Piggot, N.E., and T.L. Marsh 2004. "Does Food Safety Information Impact U.S. Meat Demand?" *American Journal of Agricultural Economics* 86(1):154-174.

Pollak, R.A., and T.J. Wales. 1981. "Demographic Variables in Demand Analysis." *Econometrica* 49(6):1533-1551.

Powell, L.M., and E. Han. 2011. "The Costs of Food at Home and Away From Home and Consumption Patterns among U.S. Adolescents." *Journal of Adolescent Health* 48:20–26.

Powell, L.M., F.J. Chaloupka, and Y. Bao. 2007. "Community, State, and Other Environmental Issues The Availability of Fast-Food and Full-Service Restaurants in the United States: Associations with Neighborhood Characteristics." *American Journal of Preventive Medicine* 33(4S):S240–S245.

Produce for Better Health Foundation. 2015. *State of the Plate: 2015 Study on America's Consumption of Fruit and Vegetables*. Retrieved from http://www.pbhfoundation.org/pdfs/about/res/pbh_res/State_of_the_Plate_2015_WEB_Bookmarked.pdf

Ramsey, J. 2010. "Bootstrapping the PC and CPC Algorithms to Improve Search Accuracy." Department of Philosophy. Technical Report No. CMU-PHIL-187, Carnegie Mellon University.

Ramsey, J.D., S.J. Hanson, and C. Glymour. 2011. "Multi-Subject Search Correctly Identifies Causal Connections and Most Causal Directions in the DCM Models of the Smith et al. Simulation Study." *NeuroImage* 58:838-848.

Ramsey, J.D., R. Sanchez-Romero, and C. Glymour. 2014. "Non-Gaussian Methods and High-Pass Filters in the Estimation of Effective Connections." *NeuroImage* 84:986-1006.

Reitzel, L.R., S.D. Regan, N. Nguyen, E.K. Cromley, L.L. Strong, D.W. Wetter, and L.H. McNeill. 2014. "Density and Proximity of Fast Food Restaurants and Body Mass Index among African Americans." *American Journal of Public Health* 104(1):110-116.

Rimm, E.B., A. Ascherio, E. Giovannucci, D. Spiegelman, M.J. Stamfer, and W.C. Willett. 1996. "Vegetable, Fruit, and Cereal Fiber Intake and Risk of Coronary Heart Disease among Men." *Journal of the American Medical Association* 275:447-451.

Rowland, C.S., J.W. Mjelde, and S. Dharmasena. 2017. "Policy Implications of Considering Pre-Commitments in U.S. Aggregate Energy Demand System." *Energy Policy* 102:406-413.

Ryan, D., and T.J. Wales. 1998. "A Simple Method for Imposing Local Curvature in Some Flexible Consumer-Demand Systems." *Journal of Business and Economic Statistics* 16(3):331-338.

SAS Institute Inc. 2014. *SAS/ETS 13.2 User's Guide*. Cary, NC: SAS Institute Inc.

Senia, M.C., H.H. Jensen, and O. Zhylyevskyy. 2017. "Time in Eating and Food Preparation among Single Adults." *Review of Economics of the Household*. 15(2):399-432.

Serletis, A., and A. Shahmoradi. 2007. "A Note on Imposing Local Curvature in Generalized Leontief Models." *Macroeconomic Dynamics* 11:290-294.

Shimizu, S., P. Hoyer, A. Hyvarinen, and A. Kerminen. 2006. "A Linear Non-Gaussian Acyclic Model for Causal Discovery." *Journal of Machine Learning Research* 7:2003-2030.

Smith, J.P. 2007. "The Impact of Socioeconomic Status on Health over the Life-Course." *Journal of Human Resources* 42(4):739-764.

Smith, T.A., J.P. Berning, X. Yang, G. Colson, and J.H. Dorfman. 2016. "The Effects of Benefit Timing and Income Fungibility on Food Purchasing Decisions among Supplemental Nutrition Assistance Program Households." *American Journal of Agricultural Economics* 98(2):564-580.

Smith, S.M., K.L. Miller, G. Salimi-Khorshidi, M. Webster, C.F. Beckmann, T.E. Nichols, J.D. Ramsey, and M.W. Woolrich. 2011. "Network Modelling Methods for FMRI." *NeuroImage* 54(2):875-891.

Spirtes, P., C.N. Glymour, and R. Scheines. 2000. *Causation, Prediction, and Search,* Cambridge, MA: MIT Press.

Spirtes, P., C. Meek, and T. Richardson. 1999. "An Algorithm for Causal Inference in the Presence of Latent Variables and Selection Bias." In *Computation, Causation, and Discovery* Menlo Park, CA: AAAI Press, pp. 211-252.

Srivastava, V.K., and D E. Giles. 1987. *Seemingly Unrelated Regression Equations Models: Estimation and Inference*. New York: Dekker.

StataCorp. 2015. *Stata 14 Base Reference Manual*. College Station, TX: Stata Press. Available at http://www.stata.com/bookstore/base-reference-manual/

Stewart, H., J. Hyman, E. Frazao, J.C. Buzby, and A. Carlson. 2011. "Can Low-income Americans Afford to Satisfy MyPryamid Fruit and Vegetable Guidelines?" *Journal of Nutrition Education and Behavior* 43(3):173-179.

Strotz, R.H. 1957. "The Empirical Implications of a Utility Tree." *Econometrica* 25(2):269-280

Taylor, R., and S.B. Villas-Boas. 2016. "A Discrete Choice Analysis of the National Household Food Acquisition and Purchase Survey (FoodAPS)." *American Journal of Agricultural Economics* 98(2):513-532.

Timperio, A., K. Ball, R. Roberts, K. Campbell, N. Andrianopoulos, and D. Crawford. 2008. "Children's Fruit and Vegetable Intake: Associations with the Neighbourhood Food Environment." *Preventive Medicine* 46:331–335.

Tobin, J. 1958. "Estimation of Relationships for Limited Dependent Variables." *Econometrica* 26(1):24-36.

Todd, J.E., L. Mancino, and B.H. Lin. 2010. *The Impact of Food Away from Home on Adult Diet Quality*. Washington DC: Economic Research Service, U.S. Department of Agriculture, Economic Research Report Number 90, February.

Todd, J.E., and B. Scharadin. 2016 *Where Households Get Food in a Typical Week: Findings From USDA's FoodAPS,* Washington DC: U.S. Department of Agriculture, Economic Research Service, Economic Information Bulletin 156, July.

Tonsor, G.T., and T.L. Marsh. 2007. "Comparing Heterogeneous Consumption in US and Japanese Meat and Fish Demand." *Agricultural Economics* 37:81–91.

U.S. Census Bureau, Population Division. 2017. Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2016. Available at https://factfinder.census.gov/bkmk/table/1.0/en/PEP/2016/PEPANNRES/0100000US|0100000US.04000|0200000US1|0200000US2|0200000US3|0200000US4.

U.S. Department of Agriculture, Economic Research Service (USDA, ERS). 2014. *Average Daily Intake of Nutrients by Food Source and Demographic Characteristics, 2007-10. Retrieved from* http://www.ers.usda.gov/data-products/food-consumption-and-nutrient-intakes.aspx

U.S. Department of Agriculture, Economic Research Service (USDA, ERS). 2015. *Nutrients (Food Energy, Nutrients, and Dietary Components). Retrieved from* http://www.ers.usda.gov/data-products/food-availability-(per-capita)-data-system/.aspx

U.S. Department of Agriculture, Economic Research Service (USDA, ERS). 2016. *National Household Food Acquisition and Purchase Survey (FoodAPS): User's Guide to Survey Design, Data Collection, and Overview of Datasets*. Washington DC: November.

U.S. Department of Agriculture, Economic Research Service (USDA,ERS). 2017a. *Food Availability (Per Capita) Data System. Retrieved from* http://ers.usda.gov/data-products/food-availability-(per-capita)-data-system.aspx.

U.S. Department of Agriculture, Economic Research Service (USDA, ERS). 2017b. *National Household Food Acquisition and Purchase Survey (FoodAPS).* http://www.ers.usda.gov/data-products/FoodAPS-national-household-food-acquisition-and-purchase-survey.aspx

U.S. Department of Health and Human Services and U.S. Department of Agriculture (HHS and USDA). *Dietary Guidelines for Americans*, *2005*, 6th Edition. Washington, DC: U.S. Government Printing Office, January.

U.S. Department of Health and Human Services and U.S. Department of Agriculture (HHS and USDA). 2010. *Dietary Guidelines for Americans*, *2010,* 7th Edition. Washington, DC: U.S. Government Printing Office, December.

U.S. Department of Health and Human Services and U.S. Department of Agriculture (HHS and USDA). 2015. *2015–2020 Dietary Guidelines for Americans,* 8th Edition. Available at http://health.gov/dietaryguidelines/2015/guidelines/.

Variyam, J.N. 2008. "Do Nutrition Labels Improve Dietary Outcomes? *Health Economics*. 17:695-708.

Variyam, J., J. Blaylock, D. Smallwood. 1996. "A Probit Latent Variable Model of Nutrition Information and Dietary Fiber Intake." *American Journal of Agricultural Economics* 78:628–639.

Ver Ploeg, M., L. Mancino, J.E. Todd, D.M. Clay, and B. Scharadin. 2015. *Where Do Americans Usually Shop for Food and How Do They Travel To Get There? Initial Findings from the National Household Food Acquisition and Purchase Survey*, Washington DC: U.S. Department of Agriculture, Economic Research Service, Economic Information Bulletin 138, March.

Volpe, R., and A. Okrent. 2012. *Assessing the Healthfulness of Consumers' Grocery Purchases*. Washington DC: U.S. Department of Agriculture, Economic Research Service, Economic Information Bulletin 102, November.

Wang, Z., and D.A. Bessler. 2006. “Price and Quantity Endogeneity in Demand Analysis: Evidence from Directed Acyclic Graphs.” *Agricultural Economics* 34:87-95.

Waterlander, W., I. Steenhuis, M. de Boer, A.J. Schuit, and J.C. Seidel. 2012. “The Effects of a 25% Discount on Fruits and Vegetables: Results of a Randomized Trial in a Three-Dimensional Web-Based Supermarket.” *International Journal of Behavioral Nutrition and Physical Activity* 9:11.

Wilde, P., and C.K. Ranney. 2000. “The Monthly Food Stamp Cycle: Shopping Frequency and Food Intake Decisions in an Endogenous Switching Regression Framework.” *American Journal of Agricultural Economics* 82:200-213.

Wilde, P., J. Llobrera, and M.V. Ploeg. 2014. “Population Density, Poverty, and Food Retail Access in the United States: An Empirical Approach.” *International Food and Agribusiness Management Review* 17(A):171-186.

Williams, D.R., S.A. Mohammed, J. Leavell, and C. Collins. 2010. “Race, Socioeconomic Status, and Health: Complexities, Ongoing Challenges, and Research Opportunities.” *Annals of the New York Academy of Science* 1186:69-101.

Wolongevicz, D.M., L. Zhu, M.J. Pencina, R.W. Kimokoti, P.K. Newby, R.B. D’Agostino, and B.E. Millen. 2010. “Diet Quality and Obesity in Women: The Framingham Nutrition Studies.” *British Journal of Nutrition* 103:1223–1229.

Wosje, K.S., P.R. Khoury, R.P. Claytor, K.A. Copeland, R.W. Hornung, S.R. Daniels, and H.J. Kalkwarf. 2010. “Dietary Patterns Associated with Fat and Bone Mass in Young Children.” *American Journal of Clinical Nutrition* 92:294–303.

Zagorsky, J.L. and P.K. Smith. 2017. “The Association between Socioeconomic Status and Adult Fast-Food Consumption in the U.S.” *Economics and Human Biology,* Available online https://doi.org/10.1016/j.ehb.2017.04.004.

Zellner, A. 1962. “An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association* 57(298):348-368.

Zhen, C., M.K. Wohlgenant, S. Karns, and P. Kaufman. 2011. “Habit Formation and Demand for Sugar-Sweetened Beverages.” *American Journal of Agricultural Economics* 93:175-193.

Zheng, Y., H.M. Kaiser. 2008. “Advertising and U.S. Nonalcoholic Beverage Demand.” *Agricultural and Resource Economics Review* 37(2):147-159.

Zheng, Z., and S.R. Henneberry. 2009. “An Analysis of Food Demand in China: A Case Study of Urban Households in Jiangsu Province.” *Applied Economic Perspectives and Policy* 31(4):873-893.

Zheng, Z., and S.R. Henneberry, 2012. “Estimating the Impacts of Rising Food Prices on Nutrient Intake in Urban China.” *China Economic Review* 23(4):1090-1103.

APPENDIX A

FIBER CONTENT ASSUMPTIONS AND ASSUMED USDA NATIONAL

NUTRIENT DATABASE FOR STANDARD REFERENCE (NDB) NUMBERS

Table A.1. Assumptions for Tortillas

Type of Tortilla	Grams of Fiber per Tortilla	NDB Number	Weight (grams) for 1 tortilla from NDB	Grams of Fiber per Ounce Tortilla
Flour	1.2	18970	49	0.69
Corn	1.5	18364	24	1.76
Whole Wheat	4	28295	41	2.76
Whole Grain	4	28295	41	2.76
Multi Grain	4	28295	41	2.76

Table A.2. Assumptions for Pasta

Type of Pasta (dry)	Grams of fiber per ounce of pasta	NDB Number (assumed type)
Plain	1.0	20120
Soy	1.1	43114 (vermicelli)
Rice	1.8	20133
Vegetable	1.2	20127 (spinach)
Whole Wheat	2.4	20124
Whole Grain	2.4	20124
Egg Noodle	1.0	20109 (enriched)

Table A.3. Assumptions for Bread

Type of Bread	Grams of Fiber per ounce of bread	NDB Number
White	0.8	18069
Wheat	1.1	18064
Rye	1.6	18060
Pumpernickel	1.8	18044
Oat/Oatmeal/Bran	1.3	18037
Potato	1.8	18971
Corn	0.7	18023
Whole Wheat	1.7	18075
Whole Grain	2.1	18035

Table A.4. Assumptions for Produce

Produce (raw, whole)	Assumed size, type, weight for item (if any)	Grams fiber/pound	Grams fiber/ounce	Grams fiber/count	NDB Number
Apples	1 medium = 182 grams	10.89	0.685	4.4	09003
Apricot		9.1			09021
Arugula	1 bunch = 227 grams		0.45	3.6	11959
Artichoke	1 medium = 128 grams		1.53	6.9	11007
Avocado	1 medium = 201 grams	30.5		13.5	09037
Banana	1 bunch = 6 medium (118 grams each)	11.8	0.74	18.6 (bunch)	09040
Blackberry		24.1	1.5		09042
Blueberry		10.9			09050
Cantaloupe	1 medium melon = 552 grams	4.1		5.0	09181
Carrots	1 bunch = 6 medium = 366 grams		0.794	10.2 (bunch)	11124
Cauliflower	1 medium = 588 grams		0.567	11.8	11135
Celery	1 bunch = 8 medium stalks = 320 grams		0.454	4.8	11143
Chard	1 bunch = 12 ounces, 240 grams		0.45	5.4	11147
Cherry	1 bag = 1.5 pounds = 680 grams	9.6		14.4	09070
Coconut	1 medium = 397 grams	41.0		35.7	12104
Corn, sweet yellow	1 medium ear = 102 grams		0.57	2.0	11167
Eggplant	1 item = 548 grams		0.85	16.4	11209
Fig	1 medium = 50 grams	13.3		1.4	09089
Garlic	10 cloves = 30 grams		0.6	1.0	11215
Grape		4.1			09132
Grapefruit, pink/red	1 item = 246 grams	7.3		3.9	09112
Honeydew	1 item = 1280 grams	3.7		10.2	09184
Kale	1 bunch = 0.25 pounds, 113 grams		1.0	4	11233
Kiwifruit, green	1 item = 69 grams	13.7		2.1	09148
Kumquat	1 item = 19 grams	29.7		1.2	09149
Lemon	1 item = 58 grams	12.8		1.6	09150
Lettuce, iceberg	1 head = 539 grams			6.5	11252
Lime	1 item = 67 grams	12.7		2.0	09159
Mango	1 item = 336 grams	7.27		5.4	09176
Mushrooms, white	1 medium = 18 grams		0.3	0.2	11260
Nectarine	1 medium = 142 grams	7.27		2.4	09191

(continued)

Table A.4. ***(continued)***

Produce (raw, whole)	Assumed size, type, weight for item (if any)	Grams fiber/pound item	Grams fiber/ounce item	Grams fiber/count item	NDB Number
Okra			0.91		11278
Onion	1 medium = 110 grams		0.5	1.9	11282
Oranges	1 item = 159 grams			7.2	09205
Papaya	1 medium = 470 grams	7.8		8.0	09226
Pea, green		23.3	1.5		11304
Peach	1 medium = 150 grams	6.9		2.2	09236
Pear	1 medium = 178 grams	14.2		5.5	09252
Pineapple	1 item = 905 grams	6.36		12.7	09266
Plum	1 item = 66 grams	6.36		0.9	09279
Pomegranate	1 item = 282 grams	18.18		11.3	09286
Pepper, sweet green	1 medium = 119 grams		0.48	2	11333
Potato	1 medium = 213 grams	10	0.62	4.7	11352
Pumpkin	1 item = 5500 grams		0.14	27.5	11422
Raspberry		29.55			09302
Radish	1 bunch = 36 grams		0.5	0.8	11429
Spinach	1 bunch = 340 grams		0.6	7.5	11457
Sprouts, Alfalfa	1 bunch = 4 ounces = 113 grams		0.54	2.17	11001
Squash, summer	1 medium = 196 grams		0.31	2.2	11641
Strawberries	1 medium = 12 grams		0.6	0.2	09316
Tangerine	1 medium = 88 grams	8.18		1.6	09218
Tomatoes	1 medium = 123 grams		0.3	1.5	11529
Watercress	1 bunch = 10 sprigs = 25 grams		0.14	0.1	11591
Watermelon	1 melon = 4518 grams	1.82		18.1	09326

APPENDIX B

TEN MOST COMMON ITEM DESCRITIPON ABBREVIATIONS FOR BREAD, PASTA, AND TORTILLA

Table B.1. Ten Highest UPC Abbreviation Frequencies and Counts for Bread

Abbreviation	Count	Percent	Assumed Meaning
BRD	31889	17.43054	Bread
F	31122	17.0113	French
BR	14452	7.899469	Brand
CTL	14449	7.897829	Private Label Brand
WHI	4643	2.537866	White
WHE	3063	1.674237	Wheat
RYE	2472	1.351196	Rye
SND	1866	1.019956	N/A
IT	1861	1.017223	Italian
S-D	1713	0.936327	Sourdough

Table B.2. Ten Highest UPC Abbreviation Frequencies and Counts for Pasta

Abbreviation	Count	Percent	Assumed Meaning
MCR	8233	13.19581	Macaroni
BR	5336	8.552516	Brand
CTL	5336	8.552516	Private Label Brand
SPG	4407	7.063519	Spaghetti
E-NDL	2716	4.353192	Egg Noodle
ELBW	1079	1.729416	Elbow
PEN	1005	1.610809	Penne
RIG	866	1.388021	Rigatoni
WW	802	1.285442	Whole Whet
RTNI	682	1.093106	Rotini

Table B.3. Ten Highest UPC Abbreviation Frequencies and Counts for Tortilla

Abbreviation	Count	Percent	Assumed Meaning
TRT	4915	18.61037	Tortilla
FLR	3129	11.84778	Flour
CRN	1874	7.095797	Corn
BR	1189	4.502083	Brand
CTL	1183	4.479364	Private Label Brand
SF	978	3.703143	N/A
R	907	3.434305	N/A
LA	612	2.317304	Large
WHT	498	1.885649	White
BUR	331	1.253313	Burrito

APPENDIX C

NUMBER OF QUARTERS EACH HOUSEHOLD REPORTS A GREATER THAN ZERO TOTAL DAILY FIBER PER CAPITA CONSUMPTION

Table C.1. Quarter with Percent Greater than Zero

Quarters Greater than Zero	Frequency	Percent	Cumulative Freq.	Cumulative Percent
2	1	0.01	1	0.01
4	2	0.02	3	0.03
6	1	0.01	4	0.04
7	4	0.04	8	0.08
9	1	0.01	9	0.09
10	1	0.01	10	0.10
12	1	0.01	11	0.11
13	2	0.02	13	0.13
14	2	0.02	15	0.15
15	1	0.01	16	0.16
16	1	0.01	17	0.17
17	2	0.02	19	0.19
18	3	0.03	22	0.22
19	5	0.05	27	0.27
20	2	0.02	29	0.29
21	2	0.02	31	0.31
22	2	0.02	33	0.33
23	1	0.01	34	0.34
24	2	0.02	36	0.36
25	5	0.05	41	0.41
26	2	0.02	43	0.43
27	5	0.05	48	0.49
28	6	0.06	54	0.55
29	4	0.04	58	0.59
30	4	0.04	62	0.63
31	8	0.08	70	0.71
32	12	0.12	82	0.83

(continued)

Table C.1. (continued)

Quarters Greater than Zero	Frequency	Percent	Cumulative Freq.	Cumulative Percent
33	13	0.13	95	0.96
34	12	0.12	107	1.08
35	24	0.24	131	1.32
36	17	0.17	148	1.50
37	19	0.19	167	1.69
38	38	0.38	205	2.07
39	44	0.44	249	2.52
40	59	0.60	308	3.11
41	77	0.78	385	3.89
42	171	1.73	556	5.62
43	476	4.81	1032	10.43
44	8864	89.57	9896	100.00

APPENDIX D

MEEK'S PC ORIENTATION RULES

The three orientation rules as described in Sprirtes et al (2000) and Meek (1995) are used to orient edges. Consider the triple X − Y − Z. In this triple, X and Y are adjacent, Y and Z are adjacent, but X and Z are not adjacent. For the variables X, Y, and Z, the sepset(X,Y) = Z if $\rho(X, Y|Z) = 0$. The orientation rules are:

1) If Y is not in the sepset of X and Z. The triple X − Y − Z is directed as X → Y ← Z.

2) If X →Y, Y and Z are adjacent, X and Z are not adjacent, and there is not arrow directed at Y, then orient Y − Z as Y → Z.

3) If there is a directed path from X to Y, and an edge between Y and Z, then direct X − Y as X → Y.

APPENDIX E

CONDTIONS UNDER WHICH THE QUANTITY CHANGE FROM GAIDS WILL BE LARGER THAN QUANTITY CHANGE FROM GAIDS GIVEN THE ELASTICITY FROM GAIDS IS MORE ELASTIC THAN THE ELASTICITY FROM AIDS

We are interested in checking when $\Delta Q_{GAIDS} \geq \Delta Q_{AIDS}$ given $\varepsilon_{\text{AIDS}} \geq \varepsilon_{GAIDS}$. Since both models face the same price increase this can also be stated as checking when $\Delta Q_{GAIDS} \geq \Delta Q_{AIDS}$ given $\%\Delta Q_{GAIDS} \geq \%\Delta Q_{AIDS}$. This means we need to check the following inequality:

$$\left(1+\frac{\%\Delta Q_{GAIDS}}{100}\right) D_{GAIDS} + P_{GAIDS} \geq \left(1+\%\frac{\%\Delta Q_{AIDS}}{100}\right) D_{AIDS}.$$

Where $\%\Delta Q_{GAIDS}$ and $\%\Delta Q_{AIDS}$ are the respective percent changes in quantity found by multiplying the elasticity by the price change, D_{GAIDS} and D_{AIDS} are the respective discretionary portions of demand for each model, and P_{GAIDS} is the level of pre-committed demand such that $P_{GAIDS} + D_{GAIDS} = D_{AIDS}$.

We move $\left(1+\frac{\%\Delta Q_{GAIDS}}{100}\right) D_{GAIDS}$ to the right side and get:

$$P_{GAIDS} \geq \left(1+\frac{\%\Delta Q_{AIDS}}{100}\right) D_{AIDS} - \left(1+\%\frac{\%\Delta Q_{GAIDS}}{100}\right) D_{GAIDS}.$$

Expanding the right hand side gives:

$$P_{GAIDS} \geq D_{AIDS} - D_{GAIDS} + \left(\frac{\%\Delta Q_{AIDS}}{100}\right) D_{AIDS} - \left(\frac{\%\Delta Q_{GAIDS}}{100}\right) D_{GAIDS}.$$

Since we assumed $P_{GAIDS} + D_{GAIDS} = D_{AIDS}$, this implies that $P_{GAIDS} = D_{AIDS} - D_{GAIDS}$.

$$P_{GAIDS} \geq P_{GAIDS} + \left(\frac{\%\Delta Q_{AIDS}}{100}\right) D_{AIDS} - \left(\frac{\%\Delta Q_{GAIDS}}{100}\right) D_{GAIDS}$$

Further simplification lead to

$$\left(\frac{\%\Delta Q_{GAIDS}}{100}\right) D_{GAIDS} \geq \left(\frac{\%\Delta Q_{AIDS}}{100}\right) D_{AIDS}$$

and

$$\frac{D_{GAIDS}}{D_{AIDS}} \geq \frac{\%\Delta Q_{AIDS}}{\%\Delta Q_{GAIDS}}.$$

Finally, this can be written as

$$\frac{D_{GAIDS}}{D_{AIDS}} \geq \frac{\varepsilon_{\text{AIDS}}}{\varepsilon_{GAIDS}}$$

because we know that each elasticity faces that same price increase ($\varepsilon_{\text{AIDS}} = \%\Delta Q_{AIDS}/\%\Delta\text{Price}$ $\varepsilon_{\text{GAIDS}} = \%\Delta Q_{GAIDS}/\%\Delta\text{P}$). Thus, the ratio of discretionary demand from GAIDS and AIDS must be greater than or equal to the ratio of elasticities. This condition can be used to check when given that the elasticity is more elastic from the GAIDS, when the resulting quantity change in the GAIDS will be large than the quantity change in the AIDS.

The following table shows this condition being check for a number of scenarios given $\varepsilon_{\text{AIDS}} = 5$ and $\varepsilon_{GAIDS} = 10$.

Table E.1. Checking Condition for $\boldsymbol{\varepsilon_{\text{AIDS}} = 5}$ and $\boldsymbol{\varepsilon_{GAIDS} = 10}$.

D_{AIDS}	D_{GAIDS}	P_{GAIDS}	ΔQ_{AIDS}	ΔQ_{GAIDS}	Condition Satisfied?
100	100	0	105	110	Yes
100	90	10	105	109	Yes
100	80	20	105	108	Yes
100	70	30	105	107	Yes
100	60	40	105	106	Yes
100	50	50	105	105	Yes
100	40	60	105	104	No
100	30	70	105	103	No
100	20	80	105	102	No
100	10	90	105	101	No
100	0	100	105	100	No